Titre: Title:	Retinal Image Quality Assessment Using Supervised Classification
Auteur: Author:	Mahnaz Fasih
Date:	2014
Type:	Mémoire ou thèse / Dissertation or Thesis
Référence: Citation:	Fasih, M. (2014). Retinal Image Quality Assessment Using Supervised Classification [Mémoire de maîtrise, École Polytechnique de Montréal]. PolyPublie. https://publications.polymtl.ca/1544/

UNIVERSITÉ DE MONTRÉAL

RETINAL IMAGE QUALITY ASSESSMENT USING SUPERVISED CLASSIFICATION

MAHNAZ FASIH

DÉPARTEMENT DE GÉNIE INFORMATIQUE ET GÉNIE LOGICIEL

ÉCOLE POLYTECHNIQUE DE MONTRÉAL

MÉMOIRE PRÉSENTÉ EN VUE DE L'OBTENTION

DU DIPLÔME DE MAÎTRISE ÈS SCIENCES APPLIQUÉES

(GÉNIE INFORMATIQUE)

AOÛT 2014

UNIVERSITÉ DE MONTRÉAL

ÉCOLE POLYTECHNIQUE DE MONTRÉAL

Ce mémoire intitulé :

RETINAL IMAGE QUALITY ASSESSMENT USING SUPERVISED CLASSIFICATION

présenté par : FASIH Mahnaz
en vue de l'obtention du diplôme de : Maîtrise ès Sciences Appliquées
a été dûment accepté par le jury d'examen constitué de :

M. QUINTERO Alejandro, Doct., président
Mme CHERIET Farida, Ph.D., membre et directrice de recherche
M. LANGLOIS J. M. Pierre, Ph.D., membre et codirecteur de recherche
M. BILODEAU Guillaume-Alexandre, Ph.D., membre

DEDICATION

To my dear husband, Behzad Rahmani

ACKNOWLEDGEMENTS

I would like to express sincere gratitude to my adviser, Professor Farida Cheriet and Professor Pierre Langlois for their patience, motivation, useful comments, remarks and engagement through the learning process of this master thesis. Their guidance helped me in all the time of research and writing of this thesis.

I would like to thank my family for their love and support throughout my study. I would like to thank my office mates at the Liv4D which we shared unforgettable memories during the study.

Also, I would like to thank Diagnos inc. for providing the required data during this research work.

A great thank you also goes to my colleague Mitchel Benovoy for his great help and comments that improve my French writing. In addition, I would like to thank my dear friends who provided helps for this study, in particular Samira Ebrahimi Kahou.

RÉSUMÉ

L’évaluation de la qualité d’images de la rétine est une étape importante dans le diagnostic automatisé des maladies de l'œil. La précision du diagnostic est dépendante de ce critère, car une mauvaise qualité de l'image peut empêcher l'observation de caractéristiques importantes de l'oeil et des manifestations pathologiques. Un algorithme robuste est donc nécessaire afin d'évaluer la qualité des images dans une base de données. Les méthodes existantes sont généralement basées sur des paramètres de qualité d'image génériques ou des critères structurels. L’évaluation de la qualité de l'image en fonction de paramètres génériques exploite la plupart des paramètres tels que les caractéristiques de contraste, netteté, éclairage, et la texture. Ces méthodes évitent généralement les procédures de segmentation des structures de l'oeil. Le grand avantage de ces méthodes est leur faible complexité de calcul par rapport à celles mises au point sur la base de critères structurels. En revanche, les systèmes basés sur des critères structurels exigent l'identification de structures anatomiques, y compris le disque optique et le réseau vasculaire de la macula. Ces méthodes ont tendance à manquer de robustesse lorsqu'elles traitent des images de mauvaise qualité. Bien que les algorithmes d'évaluation de la qualité de l'image précédemment développé montrent de bonnes performances, ils ont été mis en place sur différentes bases de données en s'appuyant sur des critères différents. Une nouvelle mesure est donc nécessaire.

Nous avons développé un algorithme d'évaluation de la qualité de l'image rétinienne basée sur l'extraction de caractéristiques génériques sans segmentation a priori de l’image. La nouveauté de cet algorithme est sa simplicité et son faible coût de calcul, ainsi que sa capacité à s’adapter sur la région d'intérêt locale. Les caractéristiques génériques sont constituées de la netteté locale, le ratio de bordure, les caractéristiques de texture et l’histogramme de couleur. La probabilité cumulée de flou métrique de détection et l'algorithme de codage de longueur variable ont été utilisés pour mesurer la netteté et les caractéristiques texturales, respectivement. Les caractéristiques extraites sont combinés pour évaluer la convenance de l'image à des fins de diagnostic. Basé sur le fait que les experts médicaux se concentrent davantage sur certaines structures anatomiques telles que la macula et le disque optique, nous avons décidé de comparer les approches globales et locales. Un « support vector machine » avec un « kernel » de fonctions

de base radiales a été utilisé comme un classificateur non linéaire afin de classer les images pour les groupes utilisables et non-utilisables. La méthodologie développée a été appliquée sur 100 images avec des tailles différentes classées par un expert médical. L'expert a évalué 50 images comme mesurable et 50 comme non-mesurable. Les résultats indiquent un très bon accord entre les prédictions de l'algorithme proposé et le jugement de l'expert médical: la sensibilité et la spécificité de l'approche locale sont respectivement 94% et 92% alors qu'il était de 87% et 89% pour l'approche globale.

ABSTRACT

Retinal image quality assessment is an important step in automated eye disease diagnosis. Diagnosis accuracy is highly dependent on the quality of retinal images, because poor image quality might prevent the observation of significant eye features and disease manifestations. A robust algorithm is therefore required in order to evaluate the quality of images in a large database. Existing algorithms for estimating the retinal image quality are generally based on generic image quality parameters or structural criteria. Image quality assessment based on generic parameters mostly exploits parameters such as contrast, sharpness, illumination, and texture features. These methods generally avoid eye structure segmentation procedures. The great advantage of these methods is their low computational complexity compared to those developed based on structural criteria. In contrast, systems based on structural criteria require identification of anatomical structures including the optic disc, macula and vascular network. These methods tend to lack in robustness when dealing with poor quality images. Although previously developed image quality assessment algorithms shows good performances, they have been implemented on different databases and relying on different criteria and a new metric is required to be developed.

We developed an algorithm for retinal image quality assessment based on the extraction of generic features without images segmentation. The novelty of this algorithm is its simplicity and low computational cost and also its capability to apply on local region of interest associated with retinal images. The generic features consisting of local sharpness, edge ratio, textural features and color histogram were measured. The cumulative probability of blur detection metric and the run-length encoding algorithm were used to measure the sharpness and textural features respectively. The extracted features are combined to evaluate the image suitability for diagnosis purposes. Based on the fact that medical experts focus more on some anatomical structures such as macula and optic disc, we decided to compare global and local approaches. A support vector machine with radial basis functions was used as a nonlinear classifier in order to classify images to gradable and ungradable groups. The developed methodology was applied on 100 images with various sizes graded by a medical expert. The expert evaluated 50 images as gradable and 50 as ungradable. The results indicate a very good agreement between the proposed algorithm’s predictions and the

medical expert's judgment: the sensitivity and specificity for the local approach are respectively 94% and 92% while it was 87% and 89% for the global approach.

CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ACRONYMS

AMD	Age-related macular degeneration
BIQI	Blind image quality index
CPBD	Cumulative probability of blur detection
DCT	Discrete cosine transformation
DR	Diabetic retinopathy
FR-IQA	Full-reference image quality assessment
FS	Feature selection
GLN	Grey level non-uniformity
HGRE	High gray-level run emphasis
HSV	Hue Saturation Value
IQA	Image Quality Assessment
ISC	Image structure clustering
JNB	Just noticeable blur
GGD	Generalized Gaussian distribution
KLD	Kullback Leibler distance
KNNC	K-nearest neighbor classifier
LDC	Linear Discriminate Classifier
LRLGE	Long run low gray level emphasis
LRE	Long run emphasis
LRHGE	Long run high gray level emphasis
LGRE	Low gray-level run emphasis

MSE	Mean Squared Error
MS-SSIM	Multi-Scale Structural SIMilarity
NN	Neural Network
NPDR	Non-proliferative diabetic retinopathy
NR-IQA	No-reference image quality assessment
NSS	Natural Scene Statistics
OCT	Optical Coherence Tomography
1DD	One disc diameter
PLS	Partial Least Squares
PSNR	Peak Signal-to-Noise Ratio
PDR	Proliferative diabetic retinopathy
QDC	Quadratic Discriminate Classifier
RR-IQA	Reduced-reference image quality assessment
RRED	Reduced-reference entropic differences
RLM	Run-length matrix
RLN	Run-length non-uniformity
RP	Run percentage
RGB	Red Green Blue
ROI	Region of interest
RBF	Radial basis function
ROC	Receiver operating characteristic
SSIM	Structural SIMilarity
SVR	Support vector regression

SVM	Support vector machine
SFFS	Sequential forward floating system
SRE	Short run emphasis
SRGE	Short run low gray level emphasis
SRHGE	Short run high gray level emphasis

LIST OF APPENDIX

CHAPTER 1 INTRODUCTION

Eye diseases affect a large number of human beings and their prevalence is expected to increase in the near future [1] [2]. Proper detection and treatment of eye diseases prevents visual loss in more than 50% of patients [3]. Retinal images have been demonstrated to be very efficient tools for disease diagnosis. For instance, medical experts use ophthalmology photos taken from the retina in order to diagnose eye and systemic diseases. Eye diseases such as Diabetic Retinopathy (DR) and Age-related Macular Degeneration (AMD) manifest themselves on the retina. They are detectable by analysis of retinal images and may be treatable if detected at an early stage. Retinal images are typically captured using digital fundus photography. Factors such as operating personnel's level of experience, patient head or eye movement and blinking can significantly influence the quality of retinal fundus images. They might reduce image quality and therefore cause several undesirable effects such as uneven illumination, poor contrast, blurriness and poor visibility. Misalignment can result in the incorrect estimation of the position of important anatomical features such as the macula and optic disc.

1.1 Motivation

Photos of poor quality might lead to incorrect automatic diagnosis of a disease and the grading of its severity. Thus, an expert must first visually classify the images as gradable or ungradable. A retinal image is deemed ungradable if it is impossible to make a reliable clinical judgment regarding the presence or the type of disease. Figure 1.1 shows examples of one gradable and three ungradable retinal images. Visual analysis of large databases of retinal images is a time consuming task that can be spread among a large number of experts. However, this may lead to inter-variability between results in the quality assessment or in the diagnosis provided by different experts. Automatic analysis of retinal images is thus a possible solution for the shortage of human experts and for reproducible clinical assessment. Moreover, in order to better understand the cause and progress of diseases, one may need to analyze many images over a long period of time. Automated retinal images quality analysis can reduce the need for human intervention and provide better conditions for more studies, hence increasing the functionality of retinal disease diagnosis [4].

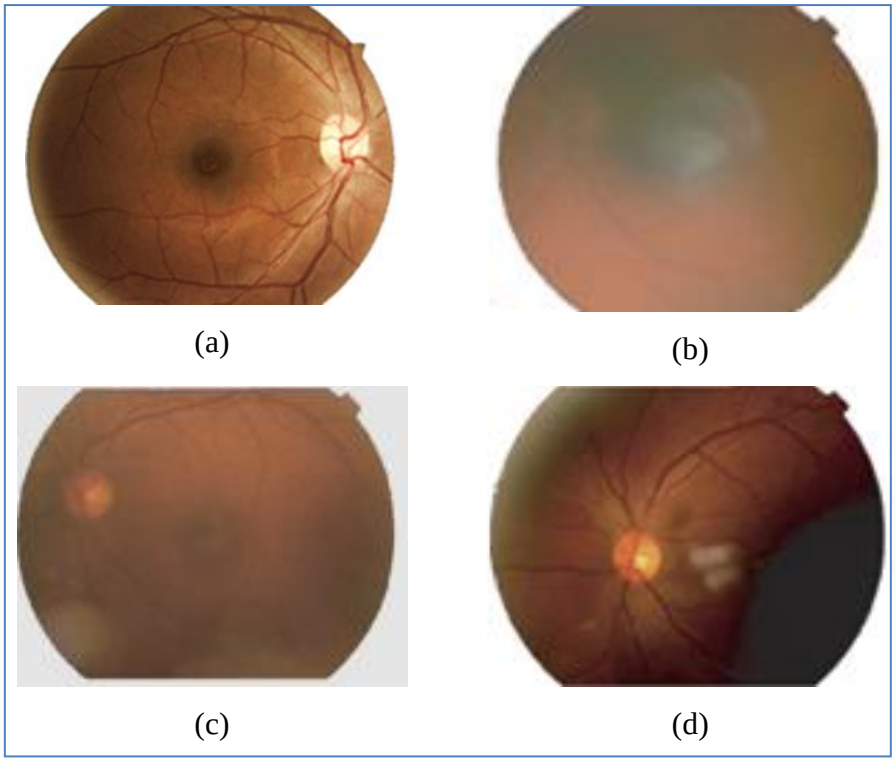


Figure 1.1: Gradable and ungradable retinal images. (a) Gradable, (b) Blur, (c) Low contrast, (d) Poor visibility of macula region and uneven illumination

1.2 Objectives

The main objective of this thesis is to propose an automatic image quality assessment approach to classify retinal images as gradable or ungradable. This general objective is divided into six specific objectives that together achieve the overall goal of the project.

i. Develop an algorithm to identify a region of interest for further analysis. Under this objective, the project will result in an algorithm to remove irrelevant image information in color retinal images that consist of a circular region (foreground) on a black background.

ii. Propose an approach to measure generic features containing textural, local sharpness, edge ratio and color histogram features. These extracted features are used for the classification of the images as gradable or ungradable.

iii. Propose a feature selection technique to reduce the dimensionality of the features vector by selecting only a subset of measured features. The goal of feature selection is to improve the effectiveness and efficiency of the classification.

iv. Construct an algorithm capable of classifying images as gradable or ungradable, according to measurements performed in the previous stages.

v. Evaluate the sensitivity and specificity of the algorithm based on experimental (visual) observations of eye experts.

vi. Evaluate the proposed method on a region including the main anatomical structures. Since experts commonly consider the macula and optic disc regions as the most useful portions of retinal images, we investigated the difference between a local and a global assessment approaches.

1.3 Thesis outline

This thesis is organized as follows. Chapter 2 introduces necessary medical background and presents information on structure of human eye, the retinal image acquisition process and retinal manifestations of systemic and eye diseases. Chapter 3 presents a literature survey and presents some recent approaches on retinal image quality assessment. Chapter 4 describes the different steps of the proposed methodology and provides some theoretical principles. Chapter 5 presents the datasets used in this study and the obtained results. These results are then discussed. The thesis ends with concluding remarks in Chapter 6.

CHAPTER 2 MEDICAL BACKGROUND AND RETINAL IMAGE ACQUISITION

This chapter presents brief reviews of the eye structure, of retinal manifestations of systemic and eye diseases, and of retinal imaging.

2.1 Eye Anatomy

The eye is made up of several structures whose collaborative action makes it possible for us to see. Several structures that compose the human eye can easily be seen by looking into a mirror (see Figure 2.1):

Figure 2.1: Front view of eye structures (waiting for permission) [5]

Iris: The iris is the colored part of the eye; it controls the amount of light entering the eye. The iris' color of each person is inherited from their parents.

Pupil: The pupil is the hole at the center of the iris that allows light to enter the eye.

Sclera: The sclera is the white part of the eye and it protects the outer layer of the eye.

The internal structure of the eye is essentially made up from a number of optical components and neural components shown in . In order to introduce these internal eye structures, let us assess the action of seeing.

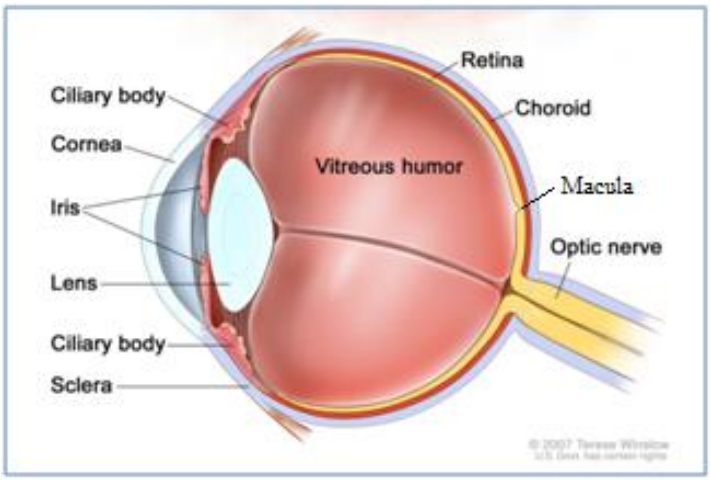


Figure 2. 2: Side view of eye structures (waiting for permission) [5]

In general in order to see, light reflects off an object and enters the eye. The light first touches a thin veil of tears that coats the front of the eye and continues its way through the cornea. The cornea helps to focus the light. On the other side of the cornea is a watery fluid called the aqueous humor. It keeps a constant pressure within the eye. Then light passes through the pupil. The iris controls the amount of light entering the eye by shrinking or widening. The light then continues through the lens. Depending on the distance of the object from the eye, the lens changes its shape in order to change the focal distance of the eye. Then light beams through the center of the eye known as the vitreous humor that surrounds the retina on which the light is focused. The retina is a light-sensitive tissue located at the back of the eye (). There are two types of nerve cells in the retina: rods and cones. The retina contains approximately 6.5 million cones and 120 million rods. The cones cells have their highest concentration in the macula. The cones cells give the brain the ability to see color and fine detail of objects while the rods cells give the brain the ability of seeing in dark places; they are not sensitive to the color of object. These two types of nerve cells convert light into electrical impulses (signals). These signals are sent from the photoreceptors of the retina to a nerve bundle that is called the optic nerve. The signals are sent by the optic nerve to the visual center in the back of the brain, allowing one to see an image. It is important to notice that the point of departure of the optic nerve through the retina does not have any rods or cones, and thus is a blind spot (optic disc). Also in the retina, there are two other important components, namely the macula and fovea (Figure 2.3). The fovea is located in the

center of the macula region. This tiny area is responsible for the sharpest vision and has a very high concentration of cones [6].

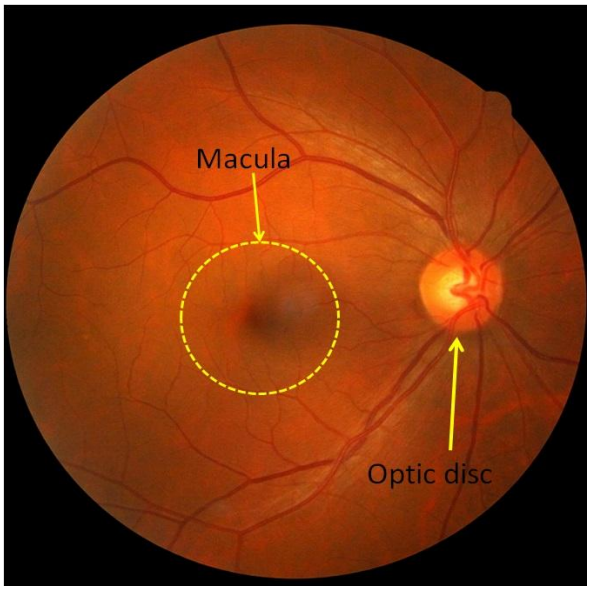


Figure 2.3: Retinal image components

2.2 Retinal Manifestations of Systemic and Eye Diseases

Retinal manifestations of systemic and eye disease are diagnosable by analyzing retinal images. These diseases are characterized through lesions which appear on the retina.

The diabetes disease can damage nerve cells and also small and large blood vessels, and consequently damage the heart, brain, kidneys and eyes, and results in a retinal complication of diabetes called diabetic retinopathy [7].

Diabetic Retinopathy: Diabetic retinopathy is the most common diabetic eye disease and a leading cause of blindness. Diabetic retinopathy can be developed in anyone who has diabetes. Diabetic retinopathy damages the blood vessels in the retina and can cause vision loss. Diabetic retinopathy is classified into two types:

I. Non-proliferative diabetic retinopathy (NPDR) is the early stage of the disease. In NPDR, the blood vessels in the retina are weakened and lead to leaking fluid into the retina, which may then lead to swelling of the macula (Figure 2.4).

II. Proliferative diabetic retinopathy (PDR) is the more advanced form of the disease. In an attempt to improve blood circulation in the retina, fragile blood vessels begin to grow in the retina. These fragile, abnormal blood vessels leads to leaking fluid into the back of the eye and block vision. PDR can cause severe vision loss and even blindness.

Figure 2.4: Effect of non-proliferative diabetic retinopathy on retina. Yellow spots are exudates and red dots are leaking blood (waiting for permission) [8]

In diabetic retinopathy, at first the vision doesn't change, but over time, diabetic retinopathy can get worse and cause vision loss [9]. Figure 2.5 shows the normal vision and vision of a person with diabetic retinopathy.

(a) (b)

Figure 2.5: Effect of diabetic retinopathy in human vision:
(a) normal vision (b) vision of person with diabetic retinopathy (permission obtained from authors) [10]

Other manifestations of eye diseases can be detected on retinal images:

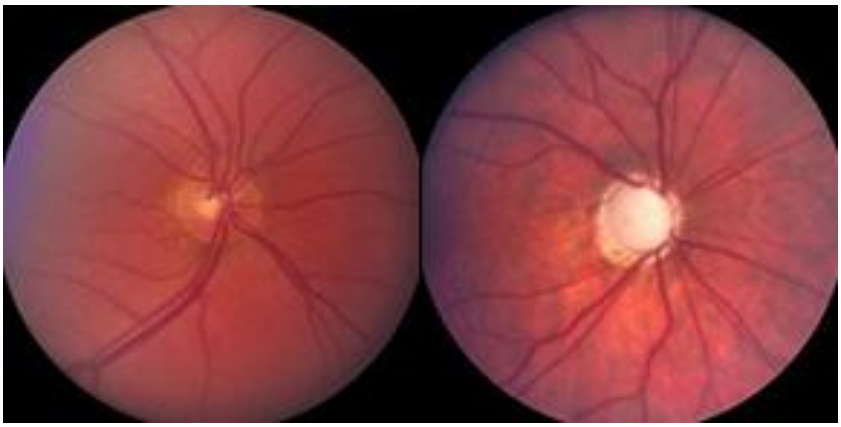
Cardiovascular Disease: Hypertension is one of the factors in cardiovascular disease. Hypertension can cause the thinning of the arteries and the widening of the veins, which increases the risk of stroke and myocardial infarction. Cardiovascular disease manifests itself in the retina such as arterial and venous occlusions, cotton wool spots and choroidal infarcts visible as deep retinal white spots. The cotton wool spots are conspicuous lesions on the retina that appear as yellow-white spots. The venous occlusion is a blockage of the small veins that carry blood away from the retina [7].

Glaucoma: Glaucoma is a condition that causes damage to the eye's optic nerve by excess fluid pressure in the eyeball, which will cause loss of vision (Figure 2.6). Without treatment, glaucoma can cause total blindness within a few years. Early diagnosis from retinal image analysis may minimize the risk of visual loss that [7, 9].

(a) (b)

Figure 2.6: Effect of Glaucoma in human vision:
(a) normal vision (b) vision of person with glaucoma (permission obtained from authors) [10]

The optic nerve head changes in the early stages of glaucoma. The cup to disc ratio is a criteria used in order to determine the risk of glaucoma. The average value for cup to disc ratio is 0.3. A large cup is a potential sign of glaucoma. The whitish area in the center of the optic disc is known as cup. Figure 2.7 shows a normal sized cup on the left and a large cup on the right damaged by glaucoma.



(a) (b)

Figure 2.7: Effect of glaucoma on the optic disc:
(a) normal optic disc, (b) optic disc damaged by glaucoma (waiting for permission) [11]

Age-Related Macular Degeneration: Age-Related Macular Degeneration (AMD) is a common eye condition and a leading cause of vision loss among people aged 50 and older. It causes damage to the macula. There are two types of AMD: dry and wet. The dry form is more common and affects about 90% of patients with AMD. The wet form is much less common, but it progresses more quickly and is more severe [9].

- Wet AMD occurs when new abnormal blood vessels start to grow in an area, such as the macula, where they are not supposed to be. These new blood vessels are very fragile and often leaking fluid that can cause damage to the macula rapidly.
- Dry AMD is an early stage of the disease. It occurs when the light-sensitive cells in the macula slowly break down. Dry AMD generally affects both eyes and patients may see a blurred spot in the center of their vision.

Figure 2.8 shows a dry form of AMD on the left and a wet form of AMD on the right.

(a) (b)

Figure 2.8: Effect of Age-Related Macular Degeneration on macula region:
(a) dry AMD, (b) wet AMD (waiting for permission) [12]

AMD leads to a blurred area near the center of vision or it may develop blank spots in the central vision, and also objects may not appear as bright as they used to be. Figure 2.9 shows the normal vision and vision of person with AMD.

(a) (b)

Figure 2.9: Effect of AMD in human vision:
(a) normal vision (b) simulated vision of person with AMD (permission obtained from authors) [10]

2.3 Retinal Imaging Modalities

There are various approaches to visualize the retina including fluorescein angiography (FA), optical coherence tomography (OCT) and fundus imaging. Each of these modalities has its own advantage and disadvantage. Ophthalmologists usually use each of these approaches in conjunction with another to provide sufficient diagnosis. Each of these modalities are described in the following.

The FA use camera and spatial dye to visualize the blood vessels of retina (Figure 2.10 (a)). In this imaging modality, a florescent dye is injected into bloodstream that fluorescein leaks from damaged retinal vessels. This can be useful for diagnosis of macular edema, diabetic retinopathy and venous occlusions. Some complications may happen when using FA such as transient nausea, vomiting and pruritis.

The OCT uses reflected infrared light to take cross-section pictures of the retina with high resolution (Figure 2.10 (b)). With OCT, each of the retina's distinctive layers can be seen. This can be useful to diagnose many conditions, including diabetic retinopathy and macular degeneration [7]. The OCT cannot illustrate entire extent of surface over which puckering may occur.

Fundus imaging still remains an important imaging modality. A fundus image is a color picture of the interior components of the eye, including the retina, optic disc and macula (as shown in Figure 2.3). The fundus images are used for diagnosis eye disease including diabetic retinopathy, glaucoma and age-related macular degeneration. Although fundus imaging is widely used, it is not suitable for fundus tomography.

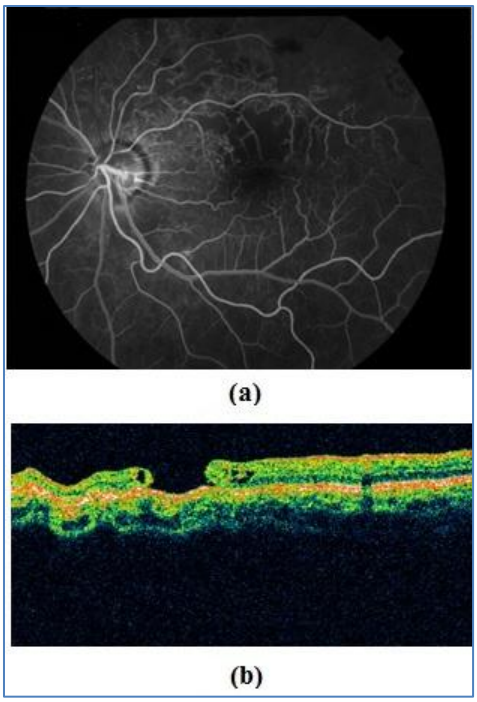


Figure 2.10 : retinal images taken by (a) fluorescein angiography [13], (b) optical coherence tomography (waiting for permission) [14]

Fundus images are taken by a digital fundus camera (Figure 2.11). The operator focuses and aligns the fundus camera and the required fundus image is created. The alignment, i.e. centering the eye horizontally and vertically, is done by moving the camera stage and high adjustment ring, respectively. The pupil must be located on the central circle on the monitor. The pupil is made to appear perfectly round by moving the camera joystick forward and backward to a position in which proper alignment is achieved. Nowadays, using modern cameras, the alignment is accomplished automatically. Retinal imaging is painless and takes a few seconds to be completed.

Figure 2.11: Fundus camera (permission obtained from authors) [10]

Fundus cameras are commonly characterized by their angle of view. A typical camera having 30° angle of view provides a magnification of 2.5×, and those with wide angle of view (between 45° and 140°) provide magnification factors between 2 × and 0.5×. A range of narrow angle of view is less than 20°, which provide relatively a higher retinal magnification. Figure 2.12 shows some examples of retina fundus obtained with different angles of view. Most of the digital cameras come with adjustable magnification. Higher magnification allows a narrow angle of view as well as a more detailed image, which is quite useful when there are more anatomical features in the shot.

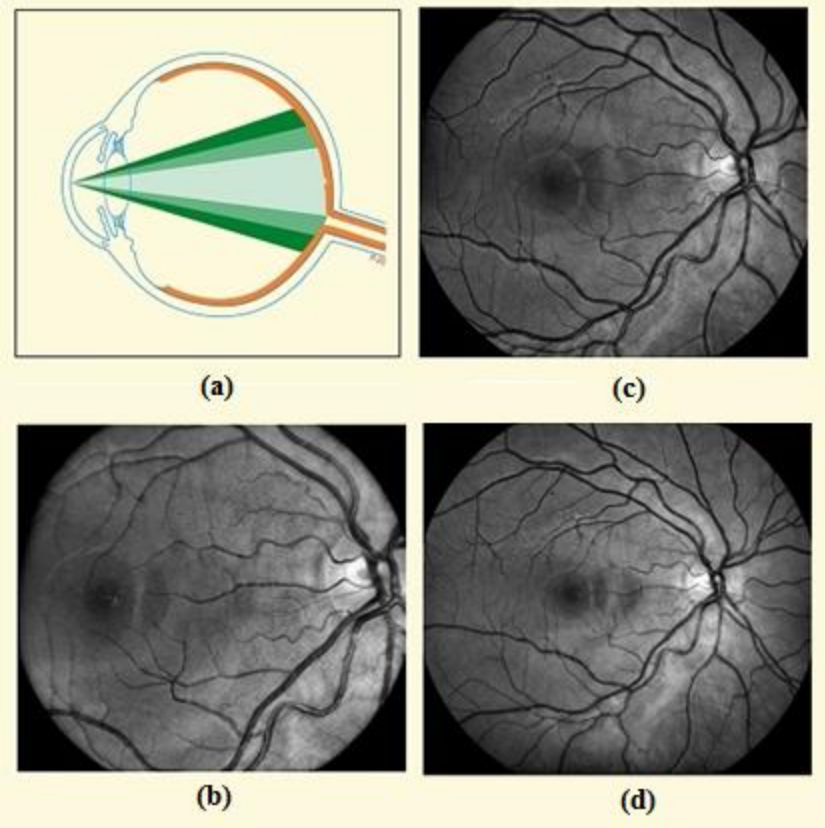


Figure 2.12: (a) A scheme of angle of view; retinal image captured by angle of view of (b) 20° (c) 40° and (d) 60° (waiting for permission) [15]

Retinal imaging can help ophthalmologist to diagnose eye diseases such as diabetic retinopathy and macular degeneration (as described in Section 2.2). The provided image may also help the physician to diagnose whether a disease is getting worse or treatment is effective. Furthermore, a digital photography can be enhanced, saved and retrieved very easily. It can be transferred electronically to a trained optical technician or retinal specialist so as to detect disease of a patient at a remote location.

For the purpose of ensuring a reliable diagnosis and a valid automated processing, one should have sufficient image quality. In fact, photos of poor quality should not be further used for diagnosis. The quality of the captured images may highly vary, since the operating personnel's level of experience differs from one person to another. There are number of factors that may cause fundus images to be classified as having low visual quality:

- **Lashes/Blink:** Patient's eye blinking during image acquisition may cause eyelashes or an eyelid to be imaged during image acquisition. Lashes or a partial blink may lead to shadow or bright reflectance in the retinal image (Figure 2.13 (a)).
- **Haze:** Haze is a white/yellow halo that appears at the edge of a retinal image and may extend toward the center of image (Figure 2.13(b)). Haze can be a consequence from incorrect camera alignment.
- **Dust:** Dirty lens may produce blurred retinal images and may lead to white dots with varying sizes on the retinal images (Figure 2.13 (c)).
- **Uneven illumination over macula/optic disc:** Retinal images are classified as ungradable images if the macula or the optic disc region is lightly/darkly shadowed or totally obscured. Figure 2.13 (d) shows an example of invisibility of macula region.
- **Blurriness:** Patient's head or eye movement might lead to out-of focus images, iris appearing in the image and uneven illumination. Also, movement of imaging system or defocus aberration may cause blurred retinal images. Figure 2.13 (e) shows an example of a blurred retinal image.
- **Arc:** Incorrect distances of patient and camera might cause arc shape to appear on the image. This arc can obscure even more than half of the retinal images (Figure 2.13 (f)).

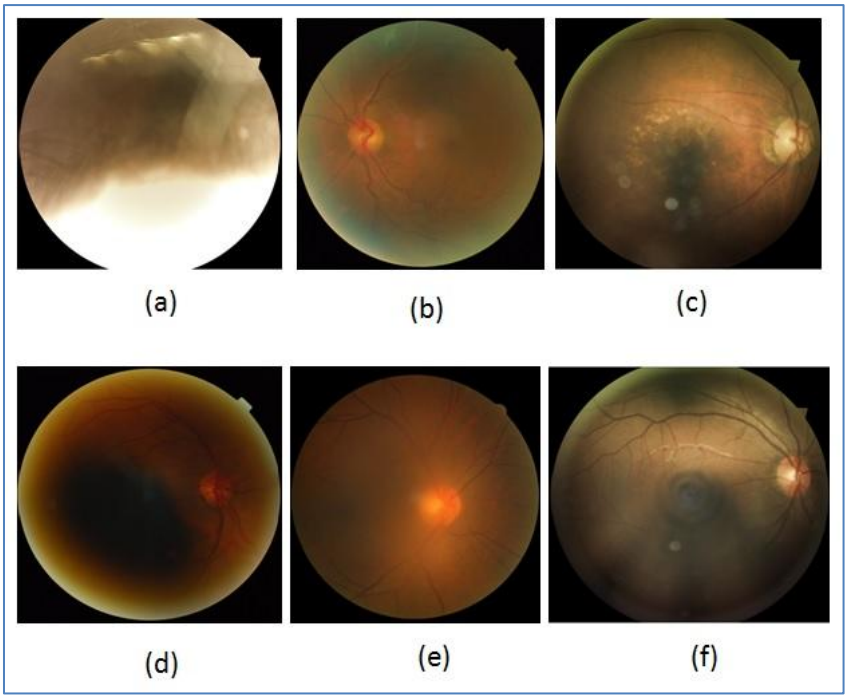


Figure 2.13 : Examples of ungradable retinal images due to (a) patient's eye blinking, (b) incorrect camera alignment, (c) dirty lens, (d) poor visibility of macula, (e) patient movement, (f) Incorrect distances of patient and camera

2.4 Conclusion

Retinal images are used for diagnose eye diseases such as diabetic retinopathy, glaucoma and macular degeneration. Automatic diagnosis must be robust enough to establish the disease's existence and to classify it. The performance of automated methods strongly depends on the quality of the images. Factors such as operating personnel's level of experience, patient head or eye movement and blinking can significantly influence the quality of retinal fundus images. Photos of poor quality might lead to incorrect diagnosis of eye disease and the grading of its severity. Thus, an expert must first classify the images as gradable or ungradable. A retinal image is deemed ungradable when it is difficult or impossible to make a reliable clinical judgment regarding the presence and type of disease it contains. Major factors that reduce image quality include uneven illumination, poor contrast, blurriness and poor visibility of important components such as the macula and optic disc. Challenge of this project is to detect these factors indicating an image of poor quality.

CHAPTER 3 LITERATURE REVIEW

This chapter presents the body of knowledge necessary to build a state-of-the-art for retinal image quality assessment, aimed at assisting automated retinal image classification.

Firstly, section 3.1 broadly describes image quality assessment techniques and their types. Section 3.2 is devoted to the state-of-art for retinal image quality assessment.

After reading this chapter, the reader should have a good idea of the state-of-the art in retinal image quality assessment, how they are built as well as the challenges to overcome in order to provide useful approaches for retinal image quality assessment.

3.1 Image quality assessment

Images can be distorted during preliminary acquisition process, compression, restoration and displaying. Image Quality Assessment (IQA) plays a significant role in several image applications. IQA in medical fields can be used in term of how well desired information can be extracted from the image. Images with acceptable quality are those that can display the image contents sufficiently. Hence, IQA is an important step before image analysis.

For several years, IQA has been performed based on subjective evaluation. In fact, The quality of fundus images is normally evaluated by human experts. That process is therefore very expensive, tedious and slow. While objective IQA evaluation compute the quality of image automatically. The goal of objective method is to evaluate the quality of an image as closely as to the human judgment. Objective IQA methods can be classified into three categories: Full-Reference IQA (FR-IQA), Reduced-Reference IQA (RR-IQA) and No-Reference IQA (NR-IQA).

3.1.1 Full-reference IQA:

The FR-IQA method evaluates the quality of an image by comparing with a reference image, which is assumed to be high quality. The simplest FR-IQA is the Mean Squared Error (MSE). The MSE is the averaged squared difference between the intensity of a reference image and a distorted image. It is computed pixel by pixel, by adding up the squared differences of all pixels intensity and dividing by the total pixel count.

For images A = $\{a_1 \dots a_N\}$ and B = $\{b_1 \dots b_N\}$, with a_i and b_i as pixels intensities:

$$MSE(A,B) = {}^{1}/_{N} \sum_{i=1}^{N} (a_i - b_i)^2 \quad (3.\ 1)$$

where N is the number of pixels.

Peak Signal-to-Noise Ratio (PSNR) is a popular FR-IQA. PSNR is the ratio between the reference signal and the distorted signal in an image, given in decibels. The higher the PSNR, the closer the distorted image is to the reference. For a reference image A = $\{a_1 \dots a_N\}$ and a distorted image B = $\{b_1 \dots b_N\}$, PSNR is calculated by:

$$PSNR(A,B) = 10\ \log_{10}\left(\frac{MAX^2}{MSE(A,B)}\right) \quad (3.\ 2)$$

where MAX is the maximum possible intensity pixel value.

The other method is the Structural SIMilarity (SSIM) index proposed by Wang *et al*. [16] which combines the luminance, contrast and structure in order to measure image quality. The SSIM between two images *x* and *y* of size N×N is calculated by the following equation:

$$SSIM(x,y) = l(x,y).c(x,y).s(x,y) \quad (3.\ 3)$$

where luminance comparison function $l(x,y)$, contrast comparison function $c(x,y)$ and structural comparison function $s(x,y)$ are defined by:

$$l(x,y) = \frac{2\bar{x}.\bar{y} + C_1}{\bar{x}^2 + \bar{y}^2 + C_1} \quad (3.\ 4)$$

$$c(x,y) = \frac{2\sigma_x.\sigma_y + C_2}{\sigma_x^2 + \sigma_y^2 + C_2} \quad (3.\ 5)$$

$$s(x,y) = \frac{\sigma_{xy} + C_3}{\sigma_x\,.\sigma_y + C_3} \quad (3.\ 6)$$

$\bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i$ and $\bar{y} = \frac{1}{N}\sum_{i=1}^{N} y_i$ are the average pixel intensity values of images *x* and *y*, $\sigma_x = \frac{1}{N-1}\sum_{i=1}^{N}(x_i - \bar{x})^2$ and $\sigma_y = \frac{1}{N-1}\sum_{i=1}^{N}(y_i - \bar{y})^2$ are the standard deviation of images *x*

and y, $\sigma_{xy} = \frac{1}{N-1}\sum_{i=1}^{N}(x_i - \bar{x})(y_i - \bar{y})$ is covariance matrix between image *x* and image *y* and C_1 , C_2 and C_3 are small constants.

Although the SSIM index is a very useful index in many experiments in some cases the measured quality does not match properly with the subjective assessment method. Wang *et al*.[17] introduced a Multi-Scale SSIM (MS-SSIM) approach in order to improve the SSIM method. In the MS-SSIM method low pass filtering and downsampling is applied iteratively and the original image is indexed as scale 1 and the finest scale *M* is obtained after $M - 1$ iterations. At each scale *j*, the given image contrast and structure is compared using $c_j(x, y)$ and $s_j(x, y)$ respectively. The luminance comparison is performed only at scale *M* and is represented as $l_M(x, y)$. The overall quality is evaluated by combining the measurements over scales. However, determining an appropriate reference for retinal images is quite difficult, since natural large variety of retinal images may be obtained from image acquisition.

3.1.2 Reduced-Reference IQA (RR-IQA):

In the RR-IQA methods the reference image is not completely available. Instead, certain features are extracted from the reference image and these features are employed to help evaluate the quality of the test image. Several methods have been proposed for RR-IQA: Wang *et al.* [18] proposed RR-IQA in which they used the wavelet transform domain based on natural image statistic model. They used the Kullback Leibler Distance (KLD) between the probability distributions of wavelet coefficients of the distorted and reference images as a measure of image quality. A Generalized Gaussian Distribution (GGD) model is employed to summarize the marginal distribution of wavelet coefficients of the reference image, so that only a relatively small number of features are needed for quality assessment. Soundarajan *et al*. [19] proposed a method for quality assessment based on Reduced-Reference Entropic Differences (RRED). This method measures differences between the entropies of wavelet coefficients of reference and distorted images for image quality evaluation. One of the remarkable features of the RRED indexes is that this method does not depend on any parameters to be trained on databases. However, similar to the FR-IQA acquiring an appropriate reference for retinal images is not simple. In addition, using features to develop a final quality score is much more difficult than FR-IQA.

3.1.3 No-reference IQA (NR- IQA):

NR-IQA or blind IQA is much more difficult to implement, because quality assessment should be performed without using any reference. Most of NR- IQA relies on learning procedures to estimate image quality from training samples, for which human subjective quality scores are available. These methods contain two step frameworks: first step is extraction of features most of which are Natural Scene Statistics (NSS) and second step is model regression in machine learning such as logistical regression, probabilistic model, Support Vector Regression (SVR) and Neural Network (NN).

Moorthy *et al*. [20] proposed blind image quality index (BIQI), in which, the generalized Gaussian distribution is used to describe the distorted image. The distorted images are then classified into five different distortion categories (noise, blur, JPEG, JPEG2000 and fast fading) using Support Vector Machine (SVM). The final image quality is then calculated by SVR for each distortion category. Saad M.A. *et al*. [21] introduced BLIINDS BLind Image Integrity Notator using Discrete cosine transformation Statistics index which is a NR-IQA method. BLIINDS extract four image statistic features from the discrete cosine transformation (DCT) coefficient. Image is first divided into blocks and generalized Gaussian distribution is fitted to each block of DCT coefficients. Model parameters retrieved from each block are pooled across space to retrieve features representative of visual quality. Probabilistic model based on multivariate generalized Gaussian density is then applied for mapping computed features to image quality. Even though NR-IQA is potentially a useful approach (since there is no need for refrences), creating an algorithm for predicting visual quality, particularly without any information about original image, is difficult.

3.2 Retinal Image Quality Assessment

Existing algorithms for estimating retinal image quality are usually NR-IQA developed relying on image histogram analysis [22, 23], generic image quality parameters [24, 25], structural criteria [26-28] or both generic and structural image quality criteria [29, 30]. Image quality assessment based on generic features mostly exploits contrast, sharpness, illumination and texture features. These methods generally avoid eye structure segmentation procedures. The great advantage of these methods is their low computational complexity compared to those developed

based on structural criteria. Systems based on structural criteria require the identification of anatomical structures including the optic disc, macula and vascular network. These methods tend to lack in robustness when dealing with poor quality images.

3.2.1 Retinal IQA with Histogram based Methods

In 1999, the first approach was presented by Lee et al. [22]. They proposed a method based on measure of similarity between a template histogram and histogram of a given retinal image. They selected a set of 20 images (from a set of 360 images) with perfect quality (as desired images) and obtained their average intensity histogram, which were referred to as template intensity histogram. The quality index Q was then obtained by convolving the template histogram with the intensity histogram of a given image. The index Q is between 0 and 1, such that Q=0 means the image is ungradable and Q=1 means the image is gradable. However, the paper not include any result. In 2001, Lalonde et al. [23] demonstrated that measuring similarity of images histograms is not sufficient for quality assessment of retinal images. In fact, they found some ungradable and gradable retinal images with similar histograms shapes. Hence, they presented a different histogram based method that was based on two criteria:

1. Distribution of the edge magnitudes in the image (d_{edge}).
2. Local distribution of pixel intensity ($d_{intensity}$).

With regard to the first criterion, they found that distributions of edge magnitude of gradable retinal images have shapes that are similar to a Rayleigh distribution with smooth fall as the intensity increases. In contrast, distribution of edge magnitude of ungradable retinal images has shapes that are similar to a Rayleigh distribution but fall rapidly as the intensity increases, as Figure 3.1 shows. They used this criterion in order to measure the focus of image, since a sharper image has higher edges magnitudes. The second criterion, inspired from histogram of Lee et al. [22], considers a region based approach instead of global approach. They segment the global histogram into uniform region by standard histogram splitting algorithm from Ohlander et al. [31].

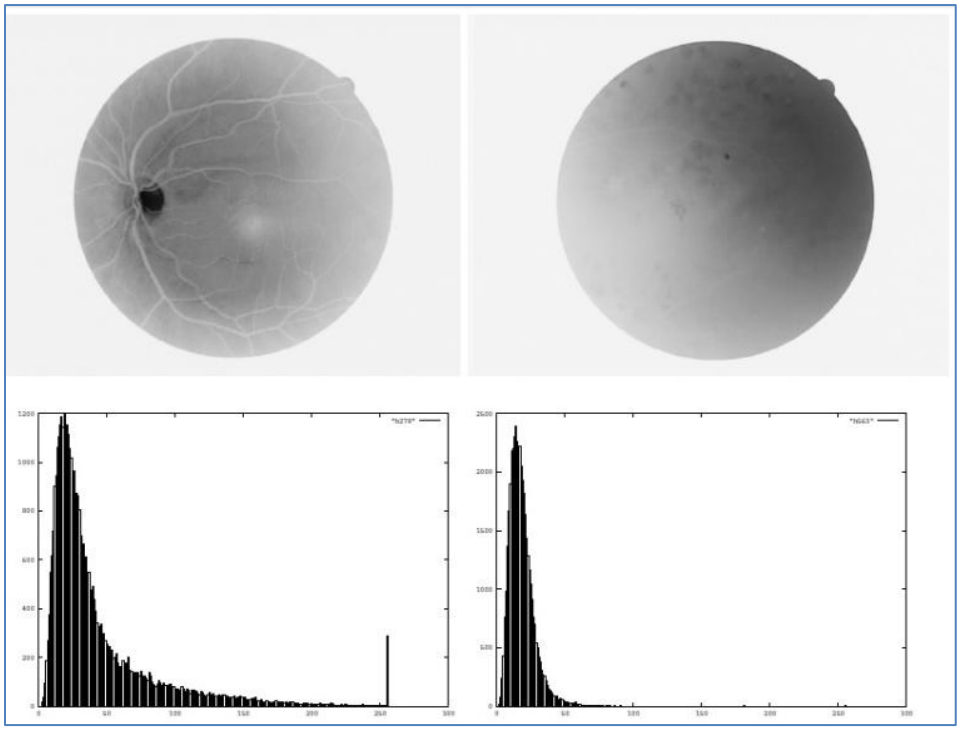


Figure 3.1: Examples of gradable (left) and ungradable (right) retinal images with the corresponding edge distributions (waiting for permission) [23]

This criterion is used in order to measure the illumination of images since gradable images should not have many dark and light pixels. By considering these two criteria, Lalonde et al. [23] tried to find the best decision border to classify images into three different categories: “Good”, “Fair” and “Bad”. Their results for 40 retinal images are shown in Figure 3.2. As shown in Figure 3.2, both criteria are able to discriminate good and bad retinal images. However, this method was implemented on only 40 retinal images that are not enough for evaluating the efficiency of the method.

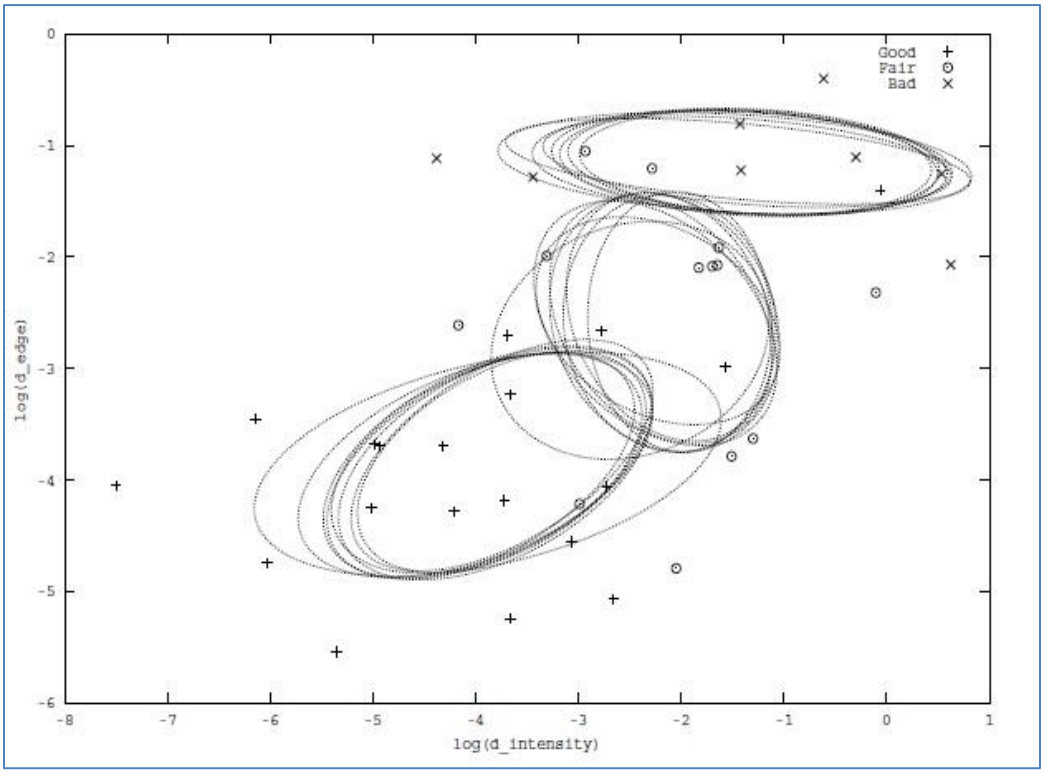


Figure 3.2: Scatter plot showing the separability of the three classes “Good”, “Fair” and “Bad” retinal images (waiting for permission) [23]

3.2.2 Retinal IQA based on Generic Image Quality Parameters

In 2009, Bartling et al. [24] developed a quality assessment method based on two generic image quality parameters: sharpness and illumination. Firstly, the images were converted to 8-bit gray scale and resized to a common resolution of 1024 pixels in width. Afterwards, the images were divided into smaller squares with 64×64 pixels, each square assessed separately.

The sharpness degree was measured on only sub-image squares having sufficient structural content. Evaluation of structural content was initiated by applying a Laplacian operator and the standard deviation of the pixel values was then calculated. The resulting value is an estimation of degree of structure content in each sub-image square. Sub-image squares with sufficient structural content were chosen based on an empirically determined threshold. These sub-image squares were then analysed in term of the spatial frequencies, since high spatial frequencies correspond to sharp image. Hence, Bartling *et al.* [24] used two dimensional discrete wavelet

transform (DWT2) using eight element Daubechies wavelet function**.** Finally, the mean value of all sub-image sharpness values was considered as the final value of the sharpness of the retinal image. Bartling *et al*. measured the image illumination through measuring image contrast and image brightness. These two parameters were measured individually in each sub-image square. They generated two weighting functions so as to measure image contrast and image brightness. Hence, contrast was calculated by using standard deviation of the pixel intensity with respect to its relevant weighting function. The image brightness was measured by using mean intensity value of each sub-image square with respect to its relevant weighting function. The final value of sub-image illumination was measured by simple addition of both measures. Sub-image squares classified by two groups (acceptable and not acceptable) based on their value of illumination. This classification was performed by an empirically determined threshold. Finally, the illumination of the whole image was determined by dividing the total number of acceptable sub-image squares by total number of sub-image squares. This value is between 0 and 1.

Score for retinal image quality was defined by the combined quality parameters (sharpness value multiplied by illumination value). Images based on their score were classified in four groups (not acceptable, acceptable, good and very good). This retinal quality measurement method was applied on 1000 retinal images and also, six independent observers classified these images into the corresponding quality groups. Hence, by comparing automatic and human quality scores, the Kappa values [32] were calculated. The median Kappa value was 0.64. The result showed that their method was not efficient enough such that the obtained results were not in agreement with medical expert judgment.

In 2009, Davis et al. [25] developed a retinal image quality assessment procedure based on seventy computationally simple features. They divided a retinal image into seven regions as shown in Figure 3.3. They converted the image from RGB color space to CIELab [33] image space. Hence, features were calculated in each color channel of CIELab space and extracted from each region. All features used by Davis et al. are listed in Table 3.1. These features were calculated in order to evaluate the color, luminance and contrast of the retinal image. They used first Quartile and third Quartile in order to evaluate the color image. They also applied mean intensity, skewness and kurtosis features in order to assess image luminance. The mean intensity feature was used for characterizing dark and light image. The skewness feature is a measure of asymmetry of the intensity histogram and kurtosis is a measure of whether the histogram of pixel

intensities is peaked or flat. Moreover, they evaluated image contrast by using variance of the intensity, generic contrast and entropy features from Haralick features and spatial frequency.

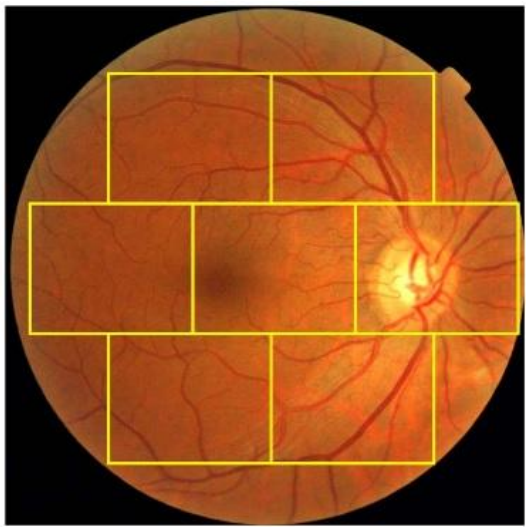


Figure 3.3: The regions where the features are calculated (waiting for permission) [25]

Table 3.1:List of features used by David et al.

Mean intensity	Skewness	Entropy	Spatial frequency	1st Quartile
Median	Variance	Kurtosis	8 Haralick features	3rd Quartile

They used variance of the intensity because of the fact that low variance tends to reflect low contrast, and the reason of using entropy is the fact that high contrast images have the largest entropy value. Partial Least Squares (PLS) was used for quality assessment, in which 398 images from the University of Iowa database [34] including 200 good quality images and 198 bad quality images were used in order to assess the procedure performance. They achieved 100% sensitivity and 96% specificity for identifying ungradable images. Satisfactory results can be obtained even though this method uses simple features. The only drawback of this method, however, is that it uses so many features (357 features) and does not use a feature selection algorithm.

3.2.3 Retinal IQA based on Structural Criteria

In 2003, Usher et al. [26] developed a first retinal image quality assessment method based on eye structures. They developed unique features for retinal image quality assessment. They noticed that higher visibility of retinal vessel implies higher retinal image quality. In fact, in a blurred image, the small vessels are not visible and also if the illumination of a retinal image is uneven, parts of the vascular network are not visible. Hence, they used a vessel segmentation algorithm within the macula region and measured the area of detected vessels. They classified images into two groups of good and bad quality by using a threshold value while this threshold value was empirically obtained. The algorithm was tested on a dataset of 1746 retinal images and it achieved 84.3% sensitivity and 95.0% specificity. The major problem of this method occurs when dealing with the distorted retinal images having some undetectable vessels outside one optic disc diameter (1DD) around the macula, while the vessels within this region are clear and sharp [35].

In 2006, Niemeijer et al. [27] developed a method based on image structure clustering (ISC). The proposed method aimed at obtaining a set of clusters using filterbanks to describe the local image structures. The features used in the final system consist of 5 bins of normalized histogram of image structure clusters in R, G and B color space. In fact, 20 features were extracted from each retinal image and features selection was employed in order to reduce the feature set. The Sequential Forward Floating System (SFFS) [36] algorithm was used for feature selection. 1000 retinal images including 500 images with normal quality and 500 images with low quality were used for the training step. They tested 1000 retinal images with different classifier methods such as SVM, Quadratic Discriminate Classifier (QDC)[37], k-Nearest Neighbor Classifier (kNNC)[37] and Linear Discriminate Classifier (LDC)[37]. The best result was obtained from SVM classifier without using feature selection algorithm. They achieved an area under the Receiver Operating Characteristic (ROC) of 0.9968 which is close to optimal. The result shows that this method can verify retinal image quality well but the authors reported a running time for an each image of approximately 30 seconds, therefore the software had not been optimized.

In 2006, Fleming et al. [28] developed a retinal image quality assessment method based on two aspects: image clarity and field definition. They calculated the total length of vessels as a measure to describe the clarity of retinal image. Therefore, they applied a vessels segmentation

algorithm within a region centered on the fovea instead of the whole image. The second aspect discussed was field definition. An image was defined as having adequate field definition if it respects all of the constraints listed in Table 3.2:

Table 3.2: Field definition metric and constraints required for adequate field definition

Field Definition Metric	Constraint
$D_{OD,Edge}$, the distance between optic disc and the edge of the circular image	$D_{OD,Edge} > 0.5\ DD$
$D_{Fovea-Edge}$, the distance between fovea and the edge of the circular image.	$D_{Fovea-Edge} > 2.0\ DD$
$\theta_{OD-Fovea}$, the angle of the line joining the center of the optic disc to the fovea	$24.7^{\circ} > \theta_{OD-Fovea} > -5.7^{\circ}$
D_{Arcade}, the length of the vessel arcades	$D_{Arcade} > 2.1\ DD$

They used disc diameter (DD) as a dimensional unit (246 pixels). Hence, for the purpose of implementing field definition aspect, retinal components segmentations are necessary. The Hough transform was applied to detect arcade vessel by employing semielliptical template. And also the Hough transform was applied to detect the optic disc with a circular template. The fovea was detected by identifying maximum cross-correlation between the image and predefined fovea template.

They employed 1039 retinal images which were graded by a medical expert for image clarity and field definition. They reported a sensitivity of 100% and specificity of 90.9% for the detection of inadequate image clarity, a sensitivity of 95.3% and specificity of 96.4% for the detection of inadequate field definition and a sensitivity of 99.1% and specificity of 89.4% for the detection of inadequate overall quality. According to the obtained results this method enables to classify retinal images in two groups: good and bad quality. In 2008, Giancardo et al. [38] presented a method inspired from the method of Niemeijer et al. [27]. They only focused on eye vessels as shown in Figure 3.4.

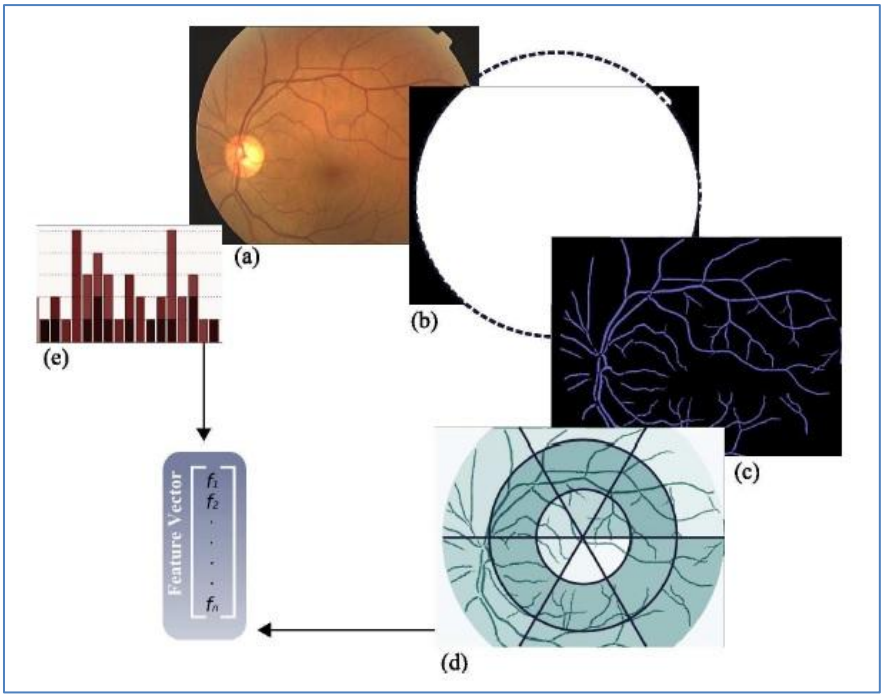


Figure 3.4: Feature vector composition reported by Giancardo et al.[38], (a) Original Image, (b) Mask detected and fitted ellipse (dashed line), (c) Vessel Segmentation, (d) Local window in polar coordinates and (e) Color histogram (waiting for permission) [38]

The mathematical morphology method presented by Zana and Klein [39] was used in order to detect the visible eye vessels. Afterward, as shown in Figure 3.4 (d), the retinal image was divided into eighteen polar windows. The vessel density of each window was computed by Equation and was stored in the feature vector.

$$Vessel\ density = \frac{Area\ vessel}{Area\ window} \tag{3. 7}$$

Moreover, 5 bins normalized histogram in RGB color spaces were stored in the feature vector. The obtained feature vector was considered as the input of a classifier algorithm. Four different classifier methods were compared: Support Vector Machines (SVM) and K-Nearest Neighbour (KNN). The employed dataset consisted of 84 macula centered fundus images from which 24 images (12 “Good” and 12 “Poor”) were extracted and used for training phase. These images

were classified by medical expert in 4 classes: “Good”, “Fair”, “Poor” and “Outliers” in order to evaluate performance of the method. Table 3.3 shows the accuracy obtained for each classifier.

Table 3.3: Performance comparison between Giancardo et al.’s method and Niemeijer et al approach

	Actual Class			
Classifier	Good	Fair	Poor	Outlier
SVM (Linear)	100%	83%	0%	11%
SVM (Radial)	100%	91%	0%	11%
KNN (k=1)	100%	66%	0%	11%
KNN (k=8)	100%	83%	0%	66%
ISC, SVM (Radial) Niemeijer et al.	83%	41%	0%	0%

These results show that Giancardo et al.’s method is suitable for good quality images. However, it cannot be used as a reliable strategy for outlier and poor quality images, since the method is based on vessel segmentation being prone to error when dealing with these kinds of images. The time consumed for quality assessment of each retinal image was around 4 seconds on a 2.4 GHz machine (with 2 GB of memory).

In 2011, Hunter et al. [40] proposed a retinal image quality assessment method based on measuring contrast and quantity of visible blood vessels. The contrast between the fovea region and retina background was calculated. The amount of contrast (κ) was considered as the first indicator of retinal image quality assessment. Hence, the contrast calculated between the average intensity of region with ten pixels radius ($r = 10$) and a doughnut-shaped region with an inner and an outer radii $r_1 = 30$ and $r_2 = 60$. The visibility of small blood vessels within one Disc Diameter (1DD) of the fovea was considered as a second indicator of quality assessment. Hence, the fovea was first located by the algorithm presented by Lowell et al. [41]. The retinal vessels were then segmented by using previously proposed algorithm of Hunter et al.[42]. Therefore, a vascular metric was defined by:

$$\nu = \sum_{i} \frac{\eta_i \alpha_i}{\gamma_i} \tag{3.8}$$

where $i \in \{1, S\}$, S is the number of vascular segment, η_i is the number of pixels in the segment, α_i is the local contrast with local background retina and $\gamma_i$is the average distance of pixels from the macula center.

The overall image quality was measured by:

$$\mu = \nu\kappa \tag{3.9}$$

These images were first classified in five groups by medical experts. Table 3. 4 shows the range of μ for five quality categories. The thresholds were empirically determined to minimize the category error using a reference set of 100 images categorized by medical expert. Hunter et al. used 200 retinal images with 760×570 pixels in order to evaluate the developed method. They achieved 100% sensitivity and 93% specificity.

Table 3. 4: Range of μ for five quality categories [40]

Category	Minimum	Maximum
1	331	-
2	101	330
3	36	100
4	6	35
5	0	5

This method focused only on the macula region, and not on the other eye components such as clarity of optic disc. Information about computation time for this methodology was not reported.

3.2.4 Retinal IQA based on Generic and Structural Criteria

In 2010, Paulus et al. [29] introduced a new approach, for evaluating retinal image quality based on both generic and structural criteria. They focused on structural information of relevant

component but without using segmentation method. They applied k-mean clustering on the pixels intensity in order to incorporate structural criteria. They chose k=5 for clustering approach empirically. Afterwards, two types of features i.e. cluster size and inter cluster differences were calculated in order to measure the recognazibility and differentiability of important components. Moreover, a sharpness metric was used to calculate the number of sharp edges and average strength of edges in order to evaluate edges' strength and detect how clearly the components are separated from each order. The sharpness metric was defined as bellow:

$$S_1 = \frac{\#\{e_{xy} | e_{xy} > \alpha\}}{n.m} \tag{3.10}$$

$$S_2 = \frac{\sum_{i=1}^{n}\sum_{j=1}^{m} v_{ij}}{\#\{e_{xy} | e_{xy} > \alpha\}}, v_{ij} = \begin{cases} 0 & e_{ij} < \alpha \\ e_{ij} & e_{ij} \geq \alpha \end{cases} \tag{3.11}$$

where S_1 is the number of sharp edges, S_2 is the average strength of edges, e_{xy} is the gradient magnitude image that are normalized into range [0;1] and *n* and *m* are the input image sizes.

To incorporate generic image quality, three Haralick features [43] were used. They used energy feature in order to describe image homogeneity, entropy feature as a description of common image sharpness and contrast feature. Five measured size of cluster, ten inter cluster differences, two sharpness metrics and three Haralick features were combined in one feature vector. Afterward, SVM with radial basis function was used as a classifier approach. The leave-one-out strategy was used to evaluate the functionality of the method. The data set used for evaluation step contained 301 retinal images (65 images with bad quality and 236 images with good quality). They tested different combination of features. The best performance was achieved when using the combination of clustering and sharpness metric and Haralick features as shown in Table 3. 5. This method is based on combination of generic and structural features and exploits the best set of generic features such as sharpness and textural features. Although this method analyzes the structural components, it does not use any segmentation algorithm. In 2012, Yu et al. [30] proposed an algorithm based on local and global features including two main steps. The first one was features extraction, in which the global histogram features, textural features and vessel density as well as local sharpness metric were extracted from retinal images. The vessel densities were measured in order to evaluate the sharpness of dark vessels because the vessel segmentation

methods are sensitive to blurry vessels. The retinal vessels were segmented using previously proposed algorithm of Yu et al. [44].

Table 3. 5: Performance of Paulus et al. method using different features [29]

Features	Sensitivity	Specificity
Haralick	90.7%	89.7%
Sharpness	46.2%	79.4%
Clustering	89.2%	86.3%
Clustering+ sharpness	87.7%	86.4%
Haralick+ sharpness	95.4%	91.0%
Haralick+ clustering	93.9%	90.4%
Haralick+ clustering+ sharpness	96.9%	91.7%

The vessel segmentation method was applied after adaptive histogram equalization and illumination correction to remove uneven illumination and enhance the contrast in retinal images. The ratio of the area of segmented vessel over the area of retinal images was calculated as the vessel density feature. Histogram features were extracted from RGB color spaces in order to describe brightness, contrast and lightness homogeneity of the image. Histogram features were variance, mean, skewness, kurtosis and the first three CDF quartiles. Five Haralick features consisting in entropy, contrast, correlation, energy and homogeneity were used in order to incorporate textural features. A local sharpness was measured by Cumulative Probability of Blur Detection (CPBD) proposed by Narvekar et al.[45]. The second step was image classification based on the extracted features. A Partial Least Square (PLS) classifier was used in order to classify images in two groups: good and bad quality. A data set consisting of 1884 retinal images (4752 ×3168) was used for evaluation step. For system validation leave-one-out cross-validation method was applied. Different feature sets and combination of all features were tested as reported in Table 3. 6. The best classification performance was achieved by a combination of all features (95.8% of AUC). The blurriness feature led to better performance than other feature sets. This method integrated global texture features, histogram features and vessel density with a local

sharpness metric. The main disadvantage of this method is that no feature reduction was applied which may lead to overfitting. The vessel segmentation algorithm was applied in order to measure vessel density, which tends to lack in robustness when dealing with poor image quality.

Table 3. 6: Results of the classifier using different features for total 1884 images

Feature sets	Area under curve (AUC)
Histogram	82.2%
Haralick	72.3%
Blurriness (including vessel density and CPBD metrics)	92.6%
All features	95.8%

3.3 Conclusion

Although all reported approaches in the literature indicate great performance scores, it is not simple to compare the obtained results of one approach to those of another. This is due to the fact that, each method may employ different database or different quality metrics for estimating their performance. In addition, it has been reported that two specialists might have different opinion regarding the quality of an image. There is no globally admitted retinal images quality assessment approach so far in the literature. Furthermore, some approach exploits either segmentation of retinal structure for quality assessment. However, since the segmentation-based approaches are error-prone, this methodology may increase the probability of error. A retinal image quality assessment approach based on generic image quality features is developed in this thesis.

CHAPTER 4 METHODOLOGY

The proposed retinal image quality evaluation method includes five processing steps. The first step is related to retinal image preprocessing. The second and third steps consist of extracting the generic features and selecting the most relevant features, respectively. In the fourth step, the selected features are fed to a Support Vector Machine (SVM) for classification. In the fifth step a performance evaluation study is carried out to compare local and global quality assessment results of the proposed method.

4.1 Preprocessing

Before analysing quality of retinal images, preprocessing operations are performed for a given image. Figure 4.1 shows the typical operations associated to preprocessing.

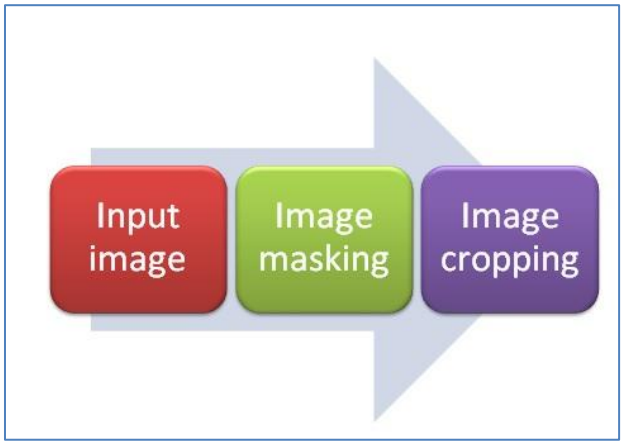


Figure 4.1: Preprocessing steps

The retinal images, consisting of different sizes and taken from different cameras, are used as input data of the preprocessing step. Image masking is used in order to label corresponding pixels of both retinal and background area throughout the image. The resulting mask is then exploited for image cropping, that is removing irrelevant image information such as useless borders. The image masking and cropping stages are detailed in the following section.

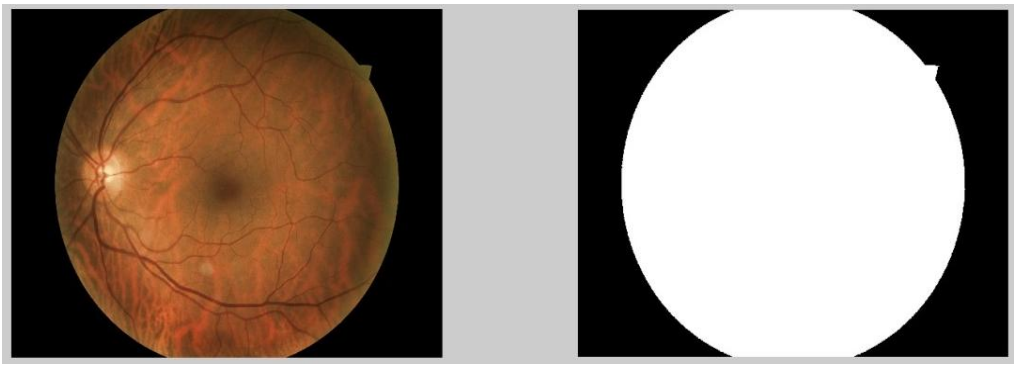
4.1.1 Image masking

A retinal image consists of a colored circular region (foreground) on a black background. It is important to distinguish background from foreground, as this preprocessing stage enables eventually to obtain more reliable results from the image quality assessment algorithm. Different algorithms have been previously developed for image masking, which are based on simple thresholding [46-48] or region growing [38].

The image masking algorithm developed in this study includes the following steps:

- Extract from a given RGB image color the red channel, as the red channel provides higher contrast between background and foreground.
- Reduce the noise by applying a median filter with mask size 3×3. The mask size was determined empirically.
- Convert the gray scale smooth image into a binary image based on an empirically determined threshold ($t = 15$).
- Apply morphological open operation on the binary image. Binary image may includ some noise at background and around the boundary of the retina. Hence, the morphological open processing is performed to removes any noises such that noise-free (sharp) edges are obtained. The kernel size for this morphological open operation is 3×3 pixels that was obtained empirically.

This results in an image for which the foreground and background pixels are specified by white and black colors, respectively, as shown in Figure 4.2 (b).



(a) (b)

Figure 4.2: (a) Orginal color retinal image and (b) masked retinal image

4.1.2 Image cropping

Cropping the retinal image reduces image size, as illustrated in Figure 4.3 (b) and therefore leads to a reduction in computation cost. Image mask created in the previous stage is used as input for the cropping process. The relative mask of each image is used in order to find the minimum bounding box around the retina. The algorithm starts searching from the left side of image mask, row by row, to find nonzero pixels which correspond to the border of retinal region. This process is repeated from the right side, to find the right border. Figure 4.3 (b) shows a typical cropped retinal image.

(a) (b)

Figure 4.3: (a) original retinal image and (b) cropped retinal image

4.2 Generic features extraction

In this step, generic features consisting of local sharpness, edge ratio, textural features and color histogram were measured. The cumulative probability of blur detection metric was used to measure the sharpness feature. The edge ratio feature was used to detect uneven illumination in the images. To incorporate texture analysis the run length [49] and co-occurrence [43] matrices were computed to determine the most efficient method for retinal texture analysis. The color information was used in order to detect overexposed or underexposed images.

4.2.1 Sharpness feature

Sharpness is an important factor for retinal image quality assessment. Retinal images might be blurred during acquisition by multiple reasons and the most common are:

- Motion: blurriness of an image due to movement of the eye, patient or imaging system.
- Lens: dirty lens setting would produce blurred retinal images.
- Defocus aberration: blurriness of a retinal image due to incorrect focus.

Figure 4.4 shows one example of sharp retinal image and three example of blurred retinal image.

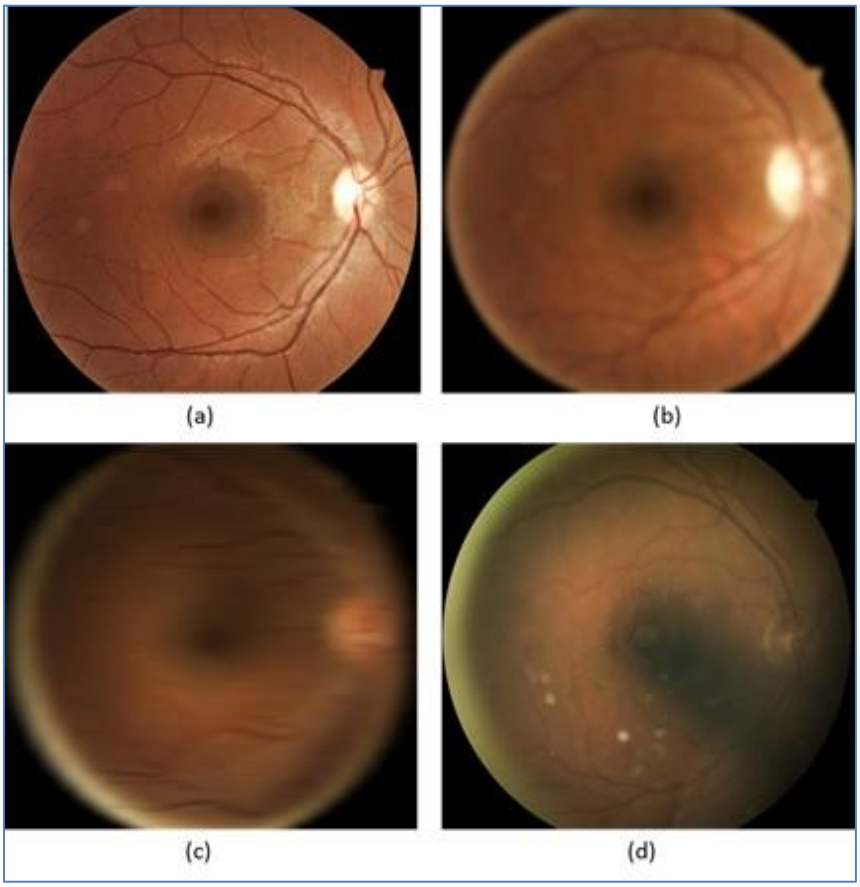


Figure 4.4: Sharp retinal image. (b) Blurred retinal image due to incorrect lens focus. (c) Blurred retinal image due to motion blur. (d) Blurred retinal image due to dirty lens

In this work, we evaluate sharpness with the Cumulative Probability of Blur Detection (CPBD) [45]. Blur estimation begins by applying edge detection. Marziliano et al. [50], tested the Sobel operator by including vertical and horizontal edges. They concluded that vertical edges did not provide any significant improvement in the results. Afterwards, the images are divided into 8×8 blocks. Depending on the edge information in each block, they are classified as edge or non-edge blocks. The blocks containing a number of edge points larger than 10% of the total number of pixels in the block are considered as edge blocks. Non edge blocks are not processed further. The rows of each edge block are scanned in order to calculate edge width $w(e_i)$ for each edge pixel (e_i). For pixels corresponding to an edge location, the start and the end positions of the un-threshold edge are defined as the locations of the local extrema close to the edge. As shown in Figure 4.5, the width of the edge is given by the distance between the end and the start positions.

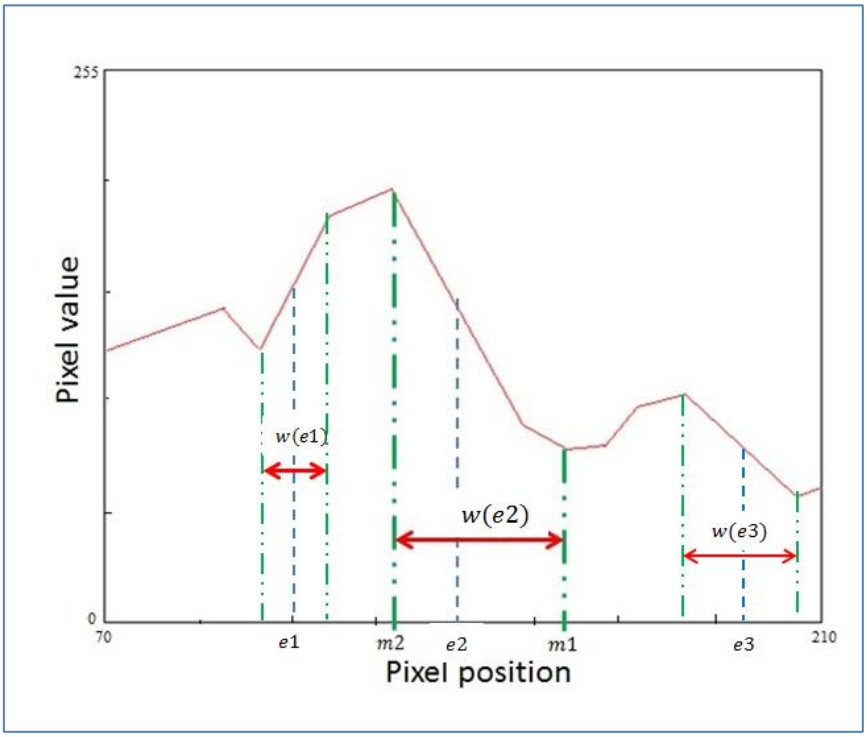


Figure 4.5: One row of the blurred image

An example of a row in an image is illustrated in Figure 4.5 for the edge location $e2$, the width is the difference between local maxima and minima around the edge $(m2 - m1)$.

The Just Noticeable Blur (JNB) edge width is obtained depending on the local contrast of the block using the following equation:

$$w_{JNB}(e_i) = \begin{cases} 5 \; if \; C \leq 50 \\ 3 \; if \; C \geq 51 \end{cases} \tag{4.1}$$

where *C* is the contrast, defined as the magnitude of the difference between the maximum and minimum intensities around the edge. The value of this parameter was obtained empirically [45] and tested for different medical images[30, 51].The probability of blur detection P_{BLUR} at each edge pixel is expressed as follows:

$$P_{BLUR} = P(e_i) = 1 - \exp\left(-\left|\frac{w(e_i)}{w_{JNB}(e_i)}\right|^{\beta}\right) \tag{4.2}$$

where the value of $\beta = 3.6$ was obtained experimentally [45], $w_{JNB}(e_i)$ is the JNB edge width and $w(e_i)$ is the measured width of the edge e_i. If the actual width of the edge is the same as the JNB edge width, then $P_{BLUR} = 63\% = P_{JNB}$. According to Narvekar and Karam [45], for images including blurred images with uniform and non-uniform saliency content and JPEG2000 compressed images, blur is not detected over an edge if $P_{BLUR} \leq 63\%$. A normalized histogram of blur detection probability is obtained, which gives the probability density function of P_{BLUR}. Finally, the Cumulative Probability of Blur Detection (CPBD) is calculated by the following equation:

$$\begin{aligned} CPBD &= P\left(P_{BLUR} \leq P_{JNB}\right) \\ &= \sum_{P_{BLUR}=0}^{P_{BLUR}=P_{JNB}} P(P_{BLUR}) \end{aligned} \tag{4.3}$$

where $P(P_{BLUR})$ denotes the value of the probability distribution function for a given P_{BLUR} [45].

4.2.2 Edge ratio feature

The edge ratio is used as another feature, since some hazy or foggy retinal images (Figure 4.6 (b)) may appear as ungradable images while the relevant edges are sharp. Figure 4.6 compares two sharp images that are classified as gradable (a) and ungradable (b) images. For this reason, the edge ratio feature is used to detect uneven illuminations. Figure 4.6 (d) and (c) also show the

corresponding edge maps. As it is shown in the figure ungradable image (d) includes much lower edges when compared with the gradable image (c).

The edge ratio began with Canny edge detection method. The Canny edge detector is an edge detection operator that uses a multi-stage algorithm to detect a wide range of edges in images [52]. The images are then divided into 8× 8 blocks. The edge ratio is the ratio of number of blocks having edge pixels more than a certain threshold to the overall number of blocks (Equation (4.4)).

$$Edge\ ratio = \frac{\text{number of edge blocks}}{\text{overall number of blocks}} \tag{4.4}$$

where, edge blocks are blocks with edge pixels larger than 10% of the total number of pixels in the block. The range of edge ratio is between 0 and 1 such that 1 represents the best quality result and 0 shows the bad quality.

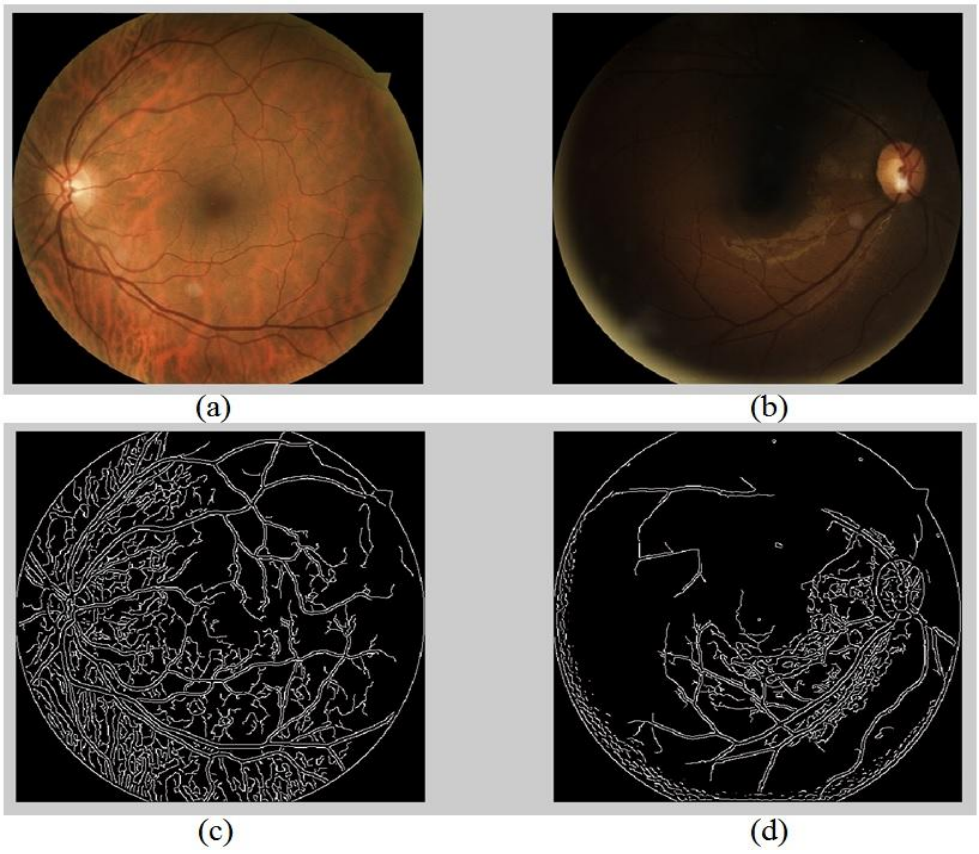


Figure 4.6: Gradable (a) and ungragable (b) images and corresponding edge maps (c and d)

4.2.3 Texture analysis

The required information in a retinal image can be extracted from its texture content. Consequently, the potential of using computerized image texture analysis for retinal image quality assessment was investigated. Selecting an appropriate feature extraction technique is a challenging aspect which can affect the quality assessment. Therefore, the two methods of feature extraction i.e. run-length matrix (RLM) [49] and co-occurrence matrix [43] were implemented and compared. Both methods are explained in details in Appendix 1.

The run-length matrix q is defined as follows. Each element $q(i, j, \theta)$ represents the number of runs with pixels of gray level intensity i and length j along the directions $\theta \in \{0°, 45°, 90°, 135°\}$.

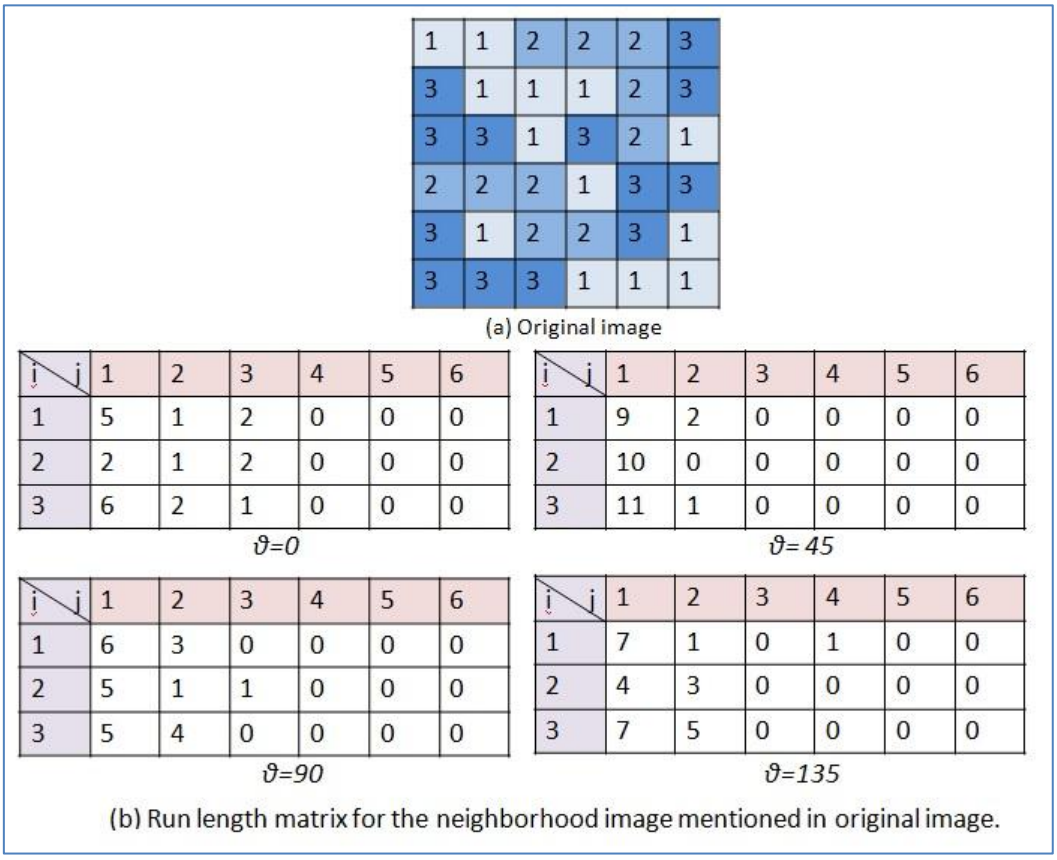


Figure 4.7: Run length matrix calculation of small neighborhood image for four different directions

Figure 4.7 shows an example of a run length matrix calculation. Figure 4.7 (a) contains a small image with 6 × 6 pixels and Figure 4.7 (b) shows the corresponding run length matrixes for four directions. For instance $q(3,2,90) = 4$ means there are 4 pixels with same intensity value of 3 with a run length of 2 for direction 90°.

In this study, the final feature was then obtained by computing the mean value of all directions. The size of the matrix q was *m* by *n*, where *m* was the maximum gray level intensity in the retinal image and *n* was equal to the possible maximum run length. Table 4.1 lists eleven considered RLM features.

Table 4.1: Run length features

Short run emphasis (SRE)	Long run low gray level emphasis (LRLGE)	Short run low gray level emphasis (SRGE)
Long run emphasis (LRE)	Long run high gray level emphasis (LRHGE)	Short run high gray level emphasis (SRHGE)
High gray-level run emphasis (HGRE)	Run-length non-uniformity (RLN)	Run percentage (RP)
Low gray-level run emphasis (LGRE)	Grey level non-uniformity (GLN)	

Features derived from a co-occurrence matrix $p(i,j)$ are effective for describing textural properties [43, 53]. Haralick [43] proposed a number of useful texture features that can be computed from the co-occurrence matrix. Table 4.2 lists 13 considered Haralick features. They were calculated with distance $d = 1$ (as described in Appendix 1) and averaged for four directions {0°, 45°, 90°, 135°}. p is an m by m matrix, where m is the maximum gray level intensity in the retinal image. Figure 4.8 shows an example for constructing a co-occurrence matrix for a small region of retinal image (5×5 pixels) with four grey levels, computed using *d=1* and $\theta = 0$.

Table 4.2: Haralick features

Energy	Sum average	Difference variance	Inverse difference moment
Contrast	Sum variance	Difference entropy	Information measure of correlation I
Correlation	Sum entropy		Information measure of correlation II
Sum of squares : variance	Entropy		

Figure 4.8 (a) contains a small region within the image (indicated by red contour) and the value of gray level intensity value associated with each cell. Figure 4.8 (b) contains the co-occurrence matrix calculated from the center pixel of the matrix (marked by a red letter) in (a). i and j in matrix (b) correspond to the intensity values. Each cell of co-occurrence matrix $p(i,j)$ contains number of times the gray level intensity values i and j have been neighbors in matrix (a). For example, assuming that d=1 and θ =0°, $p(1,0) = 3$ means that there are 3 pixels with intensity value of 1 that are in the neighborhood of pixels with 0 intensity value.

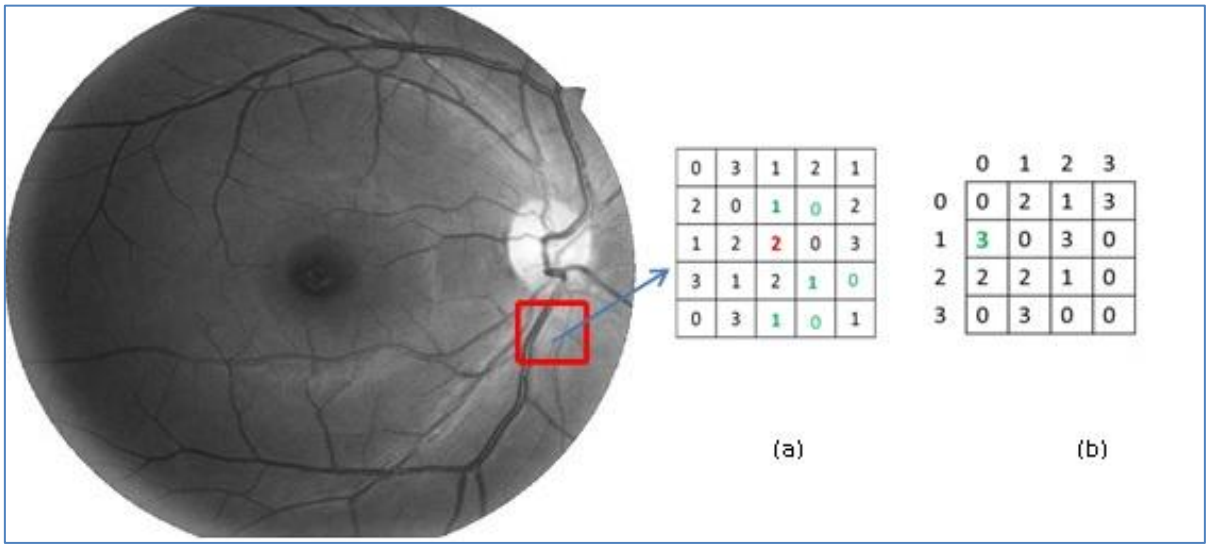


Figure 4.8: Local co-occurrence matrix calculation (a) small neighborhood image of size 5 × 5 centered at a pixel marked in red letter; (b) the corresponding co-occurrence matrix

4.2.4 Color features

We use color information to represent aspects in the quality evaluation that cannot be entirely measured by previous features. In this part the Hue Saturation Value (HSV) color space is used instead of RGB. Figure 4.9 shows HSV models.

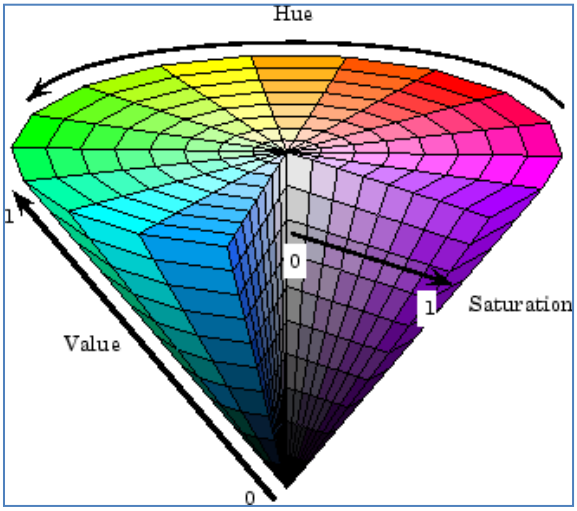


Figure 4.9: HSV color space (waiting for permission) [54]

- The Hue (H) is defined as pure spectrum color. For example all tints, tones and shades of blue have the same Hue. The values of Hue are between 0 and 360.
- The Saturation (S) describes the purity of the hue. A pure color has a saturation of 1 and tints of a color have saturations less than 1.
- The Value (V) refers to the lightness or darkness of a color. A dark color has lower amount of value (close to 0) and with increasing the lightness this value increases to 1.

The HSV color space represents the same way that humans perceive color. Colors in relation to the primary colors are defined in the other models, except for HSV. Humans can clearly categorize the colors used in HSV, which is normally different from those in RGB.

In this study, firstly, images were converted from RGB color space into HSV color space and the histogram of each layer was then obtained. The index of maximum frequency is used as a feature and is normalized between zero and one. Figure 4.10 shows three histograms of a gradable retinal image.

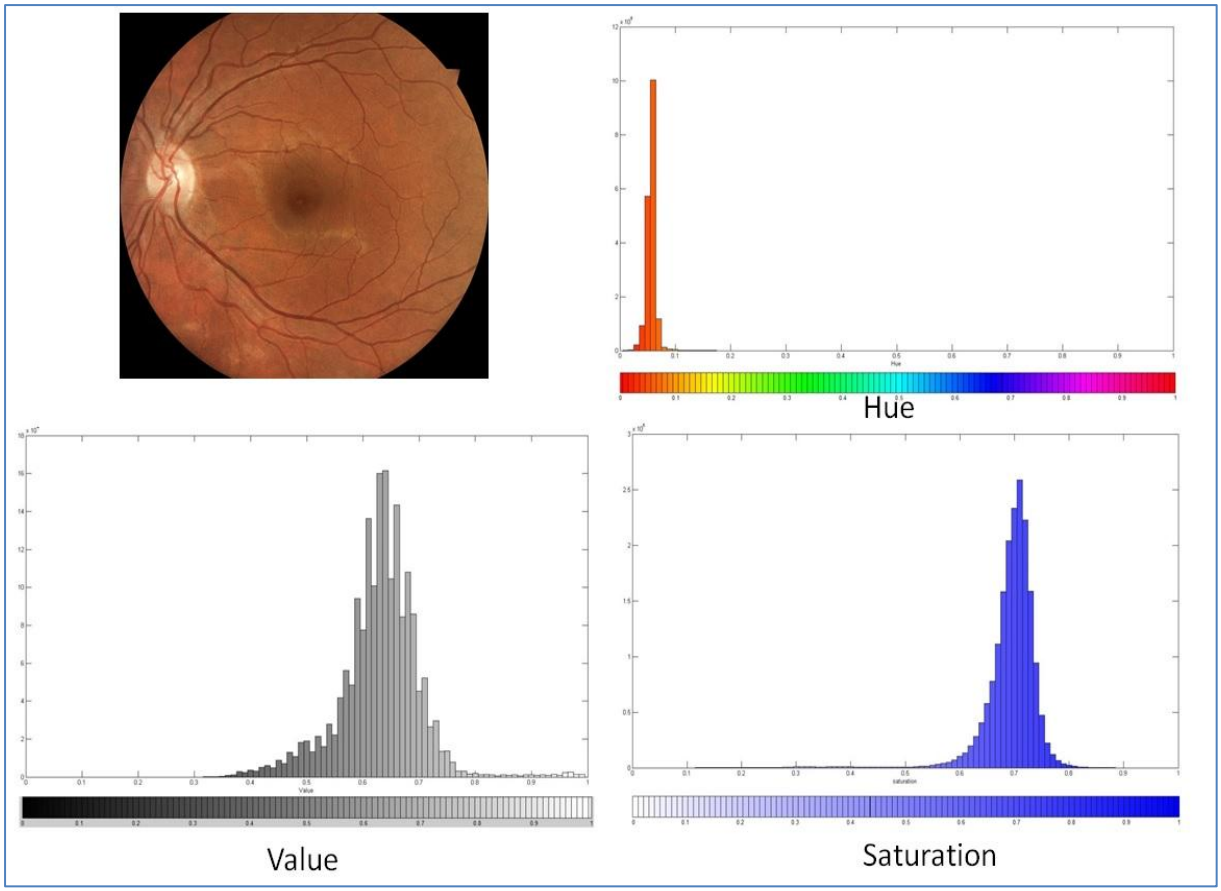


Figure 4.10: Three histograms hue, saturation and value of the gradable retinal image

The saturation and value indexes are used as features so as to detect the overexposed or underexposed images. As shown in Figure 4.11 the indexes of saturation for overexposed images ((c) and (d)) are close to zero and the indexes of value for underexposed images ((a) and (b)) are close to zero while saturation and value indexes associated with normal images ((e) and (f)) are in the range of [0.5 – 1].

The reason of not using the hue index as feature is that people with different ethnic origins have different pigmentation on the retina; Figure 4.12 shows different gradable retinal images that belong to different ethnic origins.

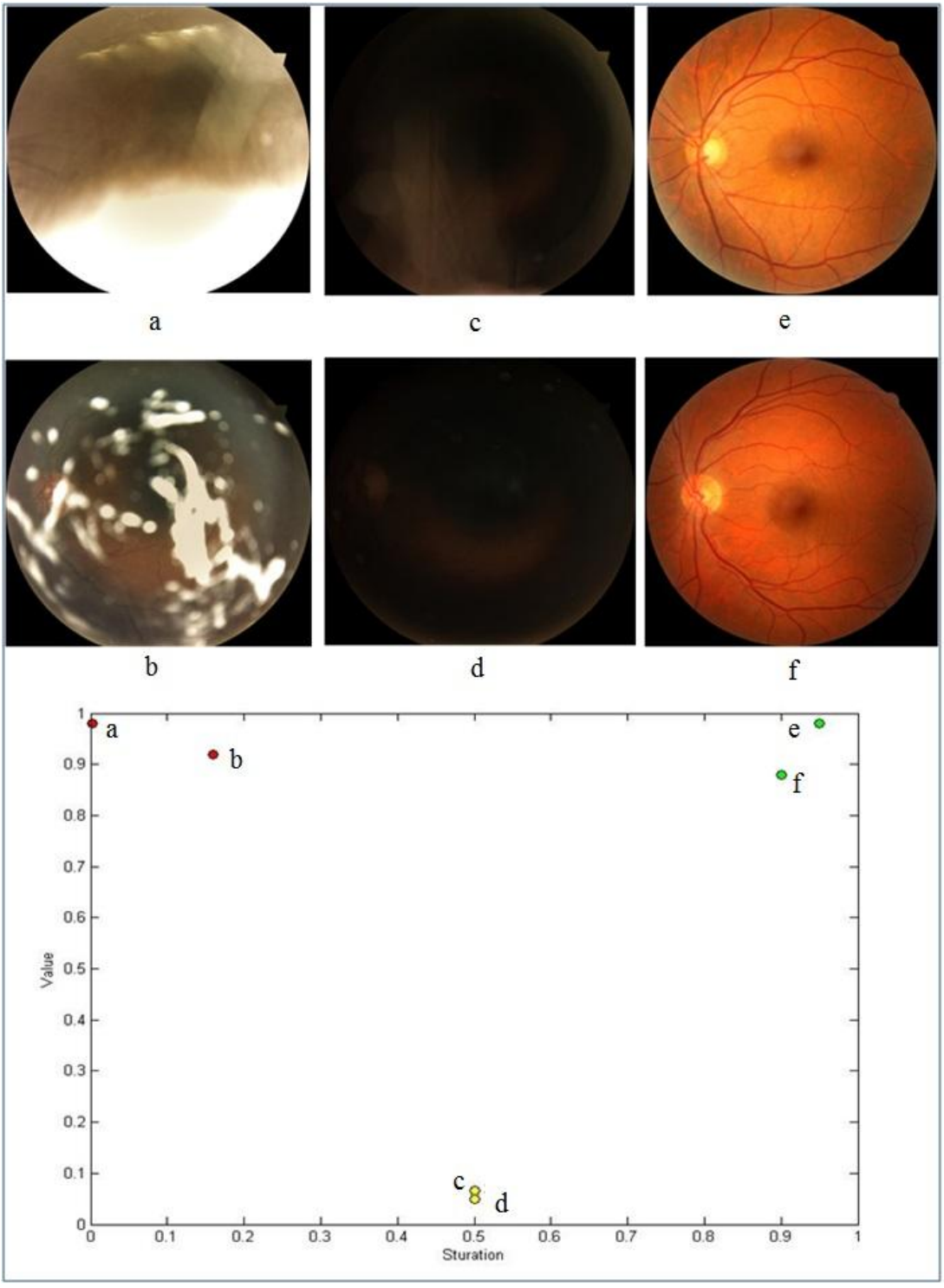


Figure 4.11: illustrate saturation and value indexes of different image quality

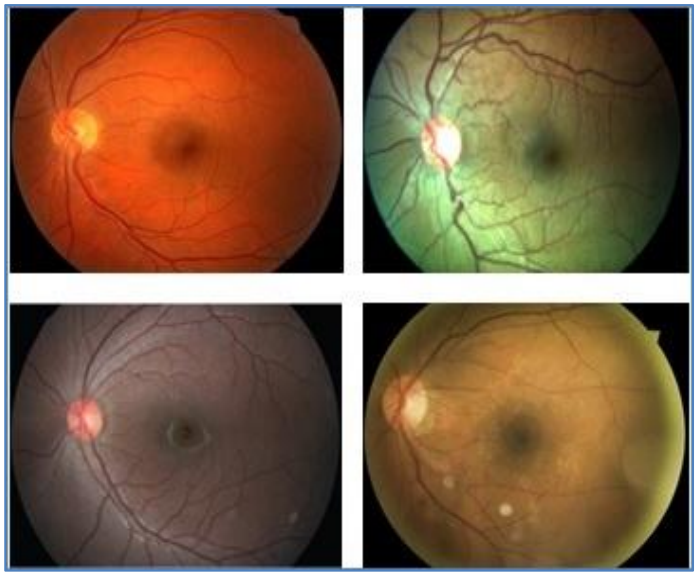


Figure 4.12: Different retinal images that belong to different ethnic origins

4.3 Feature selection

Using a large number of features may affect the classification performance. Feature selection is the process of selecting a subset of features that minimizes the classification error. We needed to perform feature selection in order to select the best set of textural features. Generally, feature selection methods are categorized into three groups: filter, wrapper and embedded methods [55]. Filter methods score the features individually and select features with high scores. Wrapper methods score feature subsets with using a predictive model. Each feature set is used to train a model, which is then tested on the remaining subset. The number of mistakes made on the test set gives the score for that subset. In the embedded methods, an optimal subset of features is searched by combining feature selection with learning algorithms. Filter methods are fast and simple and also robust against overfitting. Wrapper methods can in principle find the most useful subset of features, but are prone to overfitting. Embedded methods are embedded in and specific to a model [56]. Hence, we used filter based feature selection. There are a lot of feature selection approaches based on filter methods, such as Laplacian score [57], Fisher's score [58], ReliefF [59], Trace Ratio criterion [60] and Hilbert Schmidt Independence Criterion (HSIC) [61]. Among these, the Fisher's score is one of the most widely used because of its simplicity and good performance [62]. Hence, we used the Fisher's score [58] for sorting the features. The advantage

of this method is its simplicity and effectiveness. Fisher's score computes a rank for each feature independently, so that the rank of the j-th feature is computed using the following equation:

$$F(j) = \frac{\sum_{k=1}^{2} n_k(\mu_k^j - \mu^j)^2}{\sum_{k=1}^{2} n_k(\sigma_k^j)^2} \quad (4.5)$$

where n_k is the size of k-th class ($k \in [1,2]$) and μ^j is the mean value of the j-th feature obtained from the whole data set and μ_k^j and σ_k^j are the mean and standard deviation of k-th class associated to the j-th feature. After sorting features by means of this method, a number of appropriate features (n) should be selected based on the accuracy of the classification results. The strategy of obtaining the optimal number of features (n) is described in section 4.4.2.

4.4 Classification

As the main objective in this thesis was to classify images into two groups of gradable and un-gradable, a Support Vector Machine (SVM) with a radial basis function kernel was used as a classifier. Some common kernel functions are used for SVM include the linear kernel, the polynomial kernel and the Gaussian kernel. The linear kernel cannot handle the case may not be possible classified data linearly. The Polynomial kernel is a non-stationary kernel. It is well suited for problems where all data are normalized. The number of hyperparameters influences the complexity of model selection[63]. The polynomial kernel has more parameter than Gaussian kernel. The Gaussian kernel is known as radial basis function (RBF), RBF is one of the most popular kernel functions with convincing performance and is a reasonable to choice. Figure 4.13 shows the proposed procedure of an iteration of 10-fold cross validation used for classification. The FS process, as described in the previous section, is implemented before the classification (step 1 in Figure 4.13) for reducing the dimension of the features vector. It should be mentioned that the FS can implemented inside each cross validation iteration during classification [64]. The classification procedure consists of different steps that are detailed in the following sections.

4.4.1 Training and testing data selection

Training and testing data selection is shown as step 2 in Figure 4.13. A classification process usually involves separating training data set from testing data set. A common strategy is to separate the whole data set into two parts: the first part is considered as training data set and the

second part that is supposed to be *unknown* test data. The prediction accuracy obtained from the *unknown* data set reflects more precisely the performance of classification process, since it classifies an independent data set. An improved version of this procedure is known as k-folds cross-validation. A common choice for k-folds cross-validation is k=10. So the 10-fold cross validation technique was used for the classification in this study. As shown in Figure 4.13 (step 2), the whole data set was divided into 10 folds, such that a single fold is considered as test data (1 fold) and the remaining folds are considered as training and validation data (9 folds).

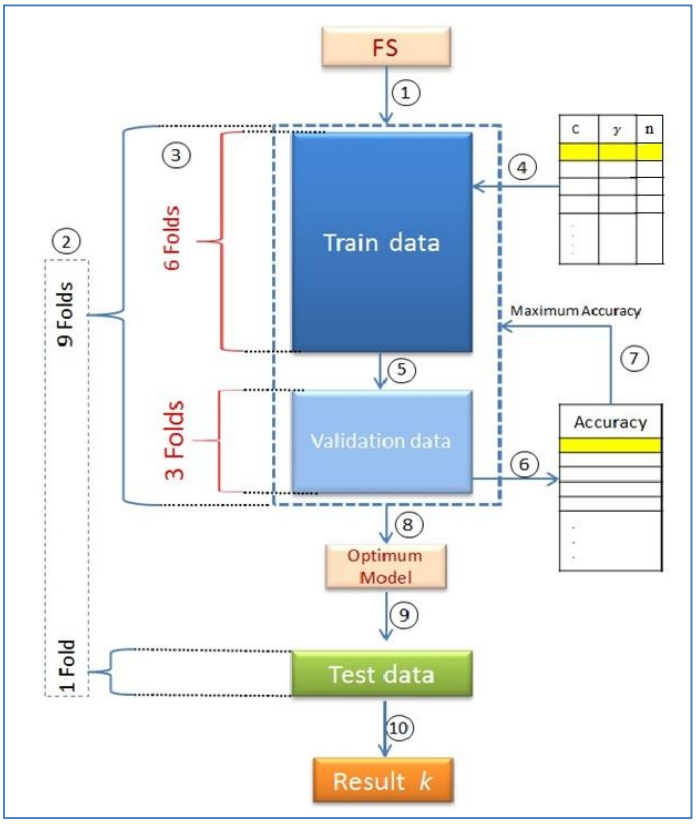


Figure 4.13: Classification steps

4.4.2 Determining optimum parameters

Steps 3-7 in Figure 4.13 include determining the optimum number of features n and kernel parameters C, γ. The features are sorted in the feature selection step, as described in the previous sections. For the purpose of reducing the feature vector dimension and particularly increasing the classification accuracy the optimum number of features should be determined. Moreover, the accuracy of an SVM model depends on the selection of the kernel parameters C, γ. We applied a

'grid-search' on *n*, *C* and γ to determine optimum values of these parameters across the specified search range. Various sets of (*n*, *C*, γ) values were tried and the one with the best cross-validation accuracy was picked. We found that trying exponentially growing sequences of C and γ is a practical method to identify good parameters (C=$10^0, 10^1, 10^2, 10^3$ and γ=$10^{-3}, 10^{-2}, \ldots, 10^2, 10^3$). In addition, the range of *n* was between 1 and the number of sorted texture features. A total of 308 combinations were totally tested. Table 4.3 shows determining optimum parameters steps.

Table 4.3: Determining optimum parameters steps

Step	The process carried out in each step
3	Divided data set into two parts (60 images for training and 30 image for validating)
4	Using a sets of (*n*, C, γ) on training data to produce SVM model
5	SVM model validation
6	Saving the result (*n*, C, γ and accuracy) from step 5
-	Repeating steps 4 to 6 until all sets of (*n*, C, γ) are evaluated
7	Finding the best *n*, C and γ corresponding to maximum accuracy

4.4.3 Evaluation

Steps 8-10 in Figure 4.13 describe the evaluation process. The optimum parameters *C* and γ were used to train the whole training set (90 images) so as to produce a model for classifying test data set. Afterward, the developed optimum model was tested on unknown test set (10 images). The accuracy of the optimum model was then calculated. As shown in Figure 4.14, in a 10-fold cross validation technique the whole data set was divided into two parts: test (1 fold) and training (9 folds) data sets. The cross-validation process is then repeated 10 times (number of folds), for which each of the 10 subsets used exactly once as the test data. So, Steps 1 to 10 in classification were repeated 10 times and the results were averaged and the final accuracy was calculated.

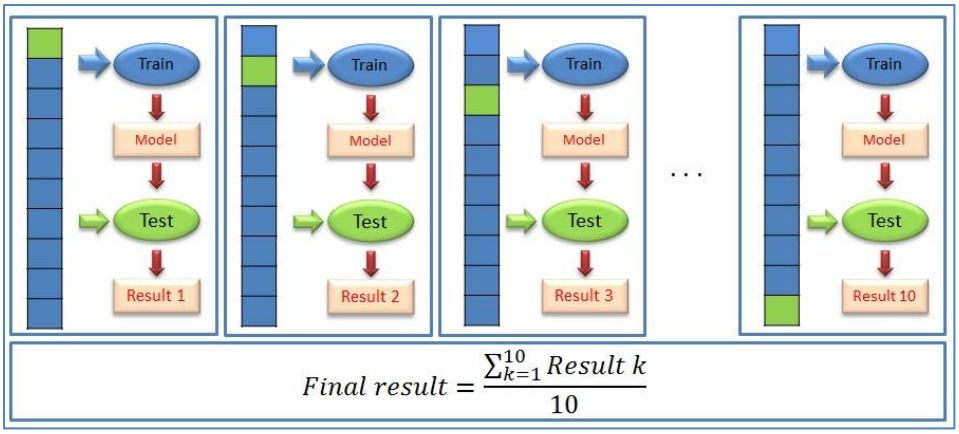


Figure 4.14: Procedure of 10-fold cross validation

4.5 Local and global evaluation

We evaluated the global and the local performance of our proposed method. In the global approach, we applied the developed algorithm to the whole image whereas in the local approach only the macula and optic disc regions are considered as regions of interest (ROI). For the purpose of localizing these two regions it was assumed that the images were centered on the macula. The input retinal image was divided into 8×8 sub-regions, through which the macula and optic disc region were determined. As shown in Figure 4.15, for a right eye, the region specified by yellow border contains the macula and optic disc.

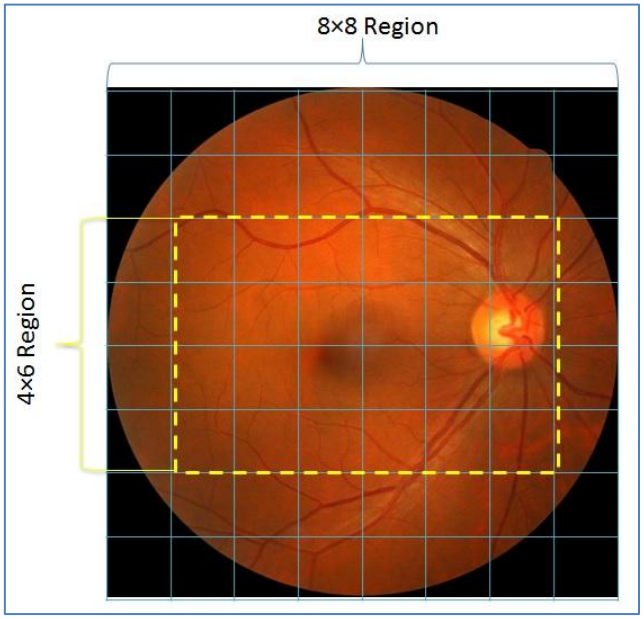


Figure 4.15 Region of interest in a typical retinal image of the right eye

CHAPTER 5 RESULTS AND DISCUSSION

This chapter presents the results associated with classification of retinal images into gradable and ungradable using different combinations of features and for global and local regions of interest.

5.1 Data set

The proposed method was evaluated using a data set consisting of 100 retinal images with different sizes. A medical expert classified 50 images as gradable and 50 images as ungradable. They were represented in RGB color space with 8 bits per color and were encoded in JPEG.

5.2 Experimental results

In this section the performance of the proposed method was assessed using a accuracy and two known statistical measures of classification i.e. sensitivity and specificity. These measures are used in order to evaluate the performance of binary classifiers, and are computed by the following equations:

$$Accuracy = \frac{tp + tn}{tp + fp + fn + tn} \quad (5.\ 1)$$

$$Sensitivity = \frac{tp}{tp + fn} \quad (5.2)$$

$$Specificity = \frac{tn}{tn + fp} \quad (5.3)$$

where tp is the number of poor quality images classified correctly, tn is the number of good quality images classified correctly, fn is the number of poor quality images classified as good quality images and fp is the number of good quality images classified as poor quality images.

5.3 Classification strategies

As described in section 4.4, the 10-folds cross-validation strategy was used in order to estimate performance of the learned model form available data using one algorithm. In other words, cross-validation used to evaluate the generalizability of the algorithm.

The database was divided into 10 folds. In each iteration, 1 fold (5 gradable images and 5 ungradable images) were considered as test set and the remaining folds were divided into two parts of 3 and 6 folds as validation and training sets, respectively. In each iteration, training and validation data sets were used so as to select the most appropriate RLM features and determine SVM kernel parameters. The test set was exploited as unseen data for evaluating the performance of the proposed method.

5.4 Texture features selection

An experiment procedure was conducted to compare the performance of the proposed method depending on using the RLM or Haralick texture features. We compared the results obtained with 11 RLM and 13 Haralick features described in section 4.2.3. The required features were extracted from the whole retinal image. In the experimentations, an SVM with Radial Basis Function (RBF) kernel was used as classifier and the results of classification were compared with those from a medical expert. The optimum SVM parameters (C and γ) that were explained in section 5.7 were used. Table 5.1 shows the results of classification obtained using the RLM and Haralick texture features.

Table 5.1 : Effectiveness of the Run Length Matrix and Haralick texture features

Features	ROI	Accuracy	Sensitivity	Specificity
Haralick texture	Whole image	74%	60%	88%
RLM	Whole image	83%	84%	82%

In this case, the texture features extracted from the RLM led to a sensitivity significantly higher than when using Haralick texture features. Hence, the RLM features were chosen for the remaining experiments.

5.5 RLM's direction selection

As described in section 4.2.3, the RLM ($f(m,n,\theta)$) is calculated in the most common directions (θ), i.e. 0, 45, 90, 135. This step aimed at finding the best direction for retinal texture analysis and therefore retinal quality assessment. Hence, each direction was taken into consideration separately and was evaluated. The SVM with RBF kernel used the optimum SVM parameters (C and γ). The accuracy of the RLM in different directions is reported in Table 5.2.

Table 5.2 : The accuracy of RLM in different directions

θ	0°	45°	90°	135°	Average of features values from four directions
Accuracy	82%	81%	80%	82%	83%

The obtained results show that, using the average of features values from four directions for retinal texture analysis, leads to higher accuracy of classification.

5.6 Feature selection

An experiment was conducted to select the most appropriate RLM features among the 11 features listed in Table 4.1. For feature selection, we applied the procedure described in section 4.3. We used a training set including 60 images (6 folds) and a validation set including 30 images (3 folds) that was selected from the database. It should be noted that in each fold there are 5 gradable and 5 ungradable retinal images. Based on the Fisher score method, a list of texture features (RLM features) was obtained. Then, in each iteration a set of features delivering higher accuracy of classification was identified. The feature selection process was repeated for each iteration of the cross validation. Table 5.3. presents the selected RLM features for 10 different iterations. In this experiment also the optimum SVM parameters were used as explained in section5.7.

Table 5.3: Selected RLM features in each iteration

Iteration	Features selected					
1	SRE	LGRE	HGRE	LRHGE	SRLGE	
2	SRE	LGRE	HGRE	LRHGE	SRLGE	
3	SRE	LGRE	HGRE	LRHGE	RP	RLN
4	SRE	LGRE	HGRE	LRHGE	SRLGE	
5	SRE	LGRE	HGRE	LRHGE	SRLGE	
6	SRE	LGRE	HGRE	LRHGE	SRLGE	
7	SRE	LGRE	HGRE	LRHGE	SRLGE	
8	SRE	LGRE	HGRE	LRHGE	RP	RLN
9	SRE	LGRE	HGRE	LRHGE	SRHGE	LRLGE
10	SRE	LGRE	HGRE	LRHGE	SRLGE	

Finally, a set of features that was dominant (i.e. was repeated more than the other sets) was selected as the best set of features. Hence, five RLM features including SRE, LGRE, HGRE, LRHGE, SRLGE were selected as best subset textures features.

5.7 SVM kernel parameters adjustment

The accuracy of an SVM is strongly dependent on the values of its parameters C and γ. A grid search strategy was used to select optimal values of these parameters in the set $\{10^{-3}, 10^{-2}, 10^{-1}, 1, 10, 10^{2}, 10^{3}\}$. For this step, the same 60 images were used as a training set and 30 images were used as validation set. We performed the classification with the RLM features, sharpness, edge ratio, color histogram features described in section 4.2. After training with the training data the classifier with the highest validation accuracy was picked for testing step. It should be noted that the optimum number of texture features n and SVM parameters C and γ were obtained at the same time. Table 5.4 reports the resulting parameters for each of the 10 iterations of the cross validation.

Table 5.4: The best obtained parameters of SVM in the each iteration

Iteration	1	2	3	4	5	6	7	8	9	10
C	10	100	1	100	10	10	10	100	1	100
γ	10	100	1	10	10	10	10	1000	1	10
n	5	5	6	5	5	5	5	6	6	5

As it can be observed, the optimum value at C and γ was 10, since this value was reported more than the other values during iteration for these parameters. The optimum number of texture features n was 5.

5.8 Classification results

The classification was performed with an SVM whose parameters were as given in Table 5.4. The complete annotated database of 100 images was used for the evaluation using a 10-fold cross validation strategy. The 10 retinal images were used as test set (unseen set) and the remaining images were used as training set. In this step, having best features (selected RLM features, one sharpness, one edge ratio and two color histogram features) and optimum parameters from the previous, the resulting SVM model in the training step was applied on a test set. The final result presents an average of the value obtained from the accuracies of ten iterations. Table 5.5 presents the classification accuracy in each iteration as well as final result obtained from average accuracy of ten iterations.

Table 5.5: Accuracy of the proposed method in each iteration

Iteration	1	2	3	4	5	6	7	8	9	10
Accuracy	80	90	90	70	90	90	90	100	70	90
Final accuracy $= \frac{80+90+90+70+90+90+90+100+70+90}{10} = 86$										

A direct comparison with results reported in the literature is difficult to make. For example [26] and [28] employed different classification i.e. good / bad quality istead of gradable / ungradable. Figure 5.1 shows example of images from our database with bad quality but that have been as garable by medical expert.

5.9 Comparison of global and local approaches

We compared the performance of the local and global approaches described in section 4.5. The same SVM parameters presented in Table 5 were used. For the global approach, the five RLM features, one sharpness, one edge ratio and two color histogram features were extracted from the whole image. In the local approach, the same six features were extracted from the most important region including the optic disc and macula. Both approaches were tested on the whole database. Table 5.6 summarizes the performance of the algorithm for the global and local approaches.

From Table 5.6, it can be concluded that the proposed method is more accurate when only the optic disc and macula regions are considered for quality assessment rather than the whole image. Analyzing portions of the image outside of the mentioned region might adversely affect the efficiency of our algorithm. Hence, the estimations can be more accurate when ignoring this region and focusing only on the macula and optic regions. Accordingly, the predictions are in very good agreement with the medical expert. I believed that, the results indicate that the proposed method is robust enough to classify retinal images according to their quality.

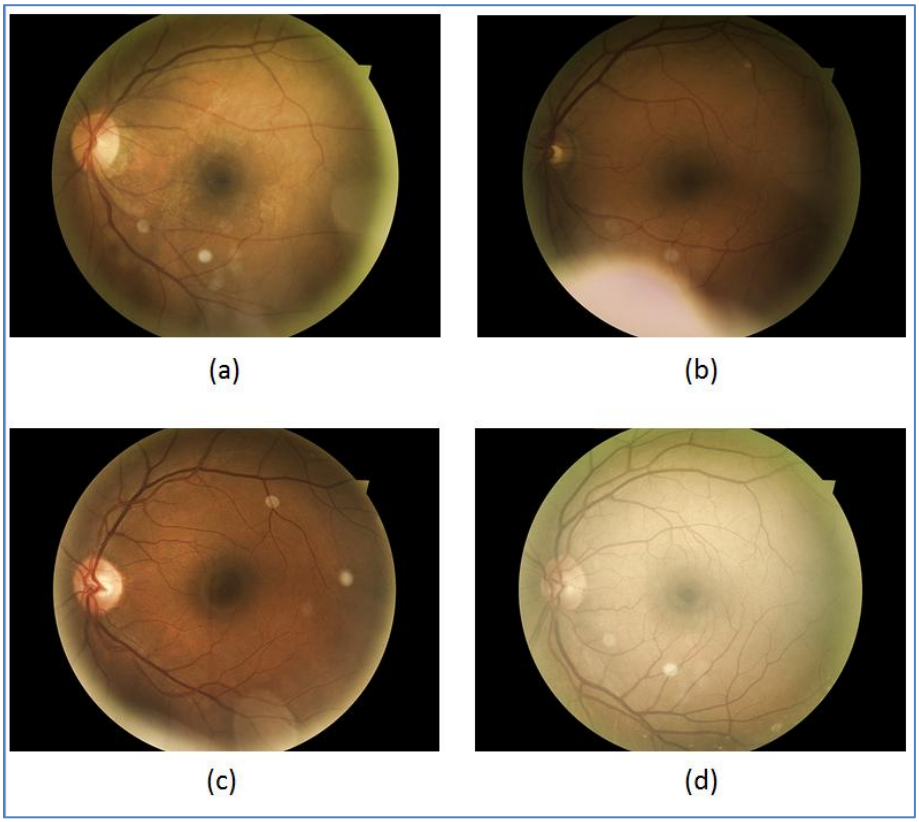


Figure 5.1: Examples of bad quality images but classified as gradable by medical expert

Table 5.6: Accuracy, sensitivity and specificity of classification resulting from the global and local approaches

ROI	Accuracy	Sensitivity	Specificity
Whole image	86%	87%	89%
Optic disc and macula regions	**93%**	**94%**	**92%**

5.10 Features classification performance

The performance of each feature was calculated. Five cases were considered: the five selected RLM features, sharpness, edge ratio and color histogram features separately and also a combination of the all features. In this experiment the SVM with RBF kernel function was used and 60, 30 and 10 retinal images were used as training, validation and testing set respectively.

Table 5.7 : Sensitivity and specificity of each class of features

Features	Accuracy	Sensitivity	Specificity
Five RLM selected features	85%	82%	88%
Sharpness	73%	60%	86%
Edge ratio	71%	70%	72%
Color histogram	81%	72%	90%
Sharpness + Edge ratio	74%	62%	86%
Sharpness + Edge ratio + color	90%	88%	92%
Combination of all features	**93%**	**94%**	**92%**

The accuracy, sensitivity and specificity of classification obtained from different combinations of features are presented in Table 5.7. As it can be seen from the results, the five RLM features and color histogram features are efficient features when considered solely. However, the combination of sharpness, edge ratio and color leads to more accurate results. Finally, the most satisfactory results were obtained when combination of all features is used in classification.

5.11 Computational efficiency of algorithm

The computation time is an important factor and it should be as low as possible. We measured processing time of the proposed algorithm for each retinal image size. The time consumed for training was not taken into account, because this is typically performed offline. The proposed algorithm was implemented by Matlab® and the tests were performed on a desktop computer with an Intel® core ™2 CPU 6600@2.4 GHz and equipped with 4GB of RAM. In this experiment, we used 5 retinal images with different resolution classified based on the optimum parameters found from previous experiment. Figure 5.2 shows the corresponding processing time for different size of images.

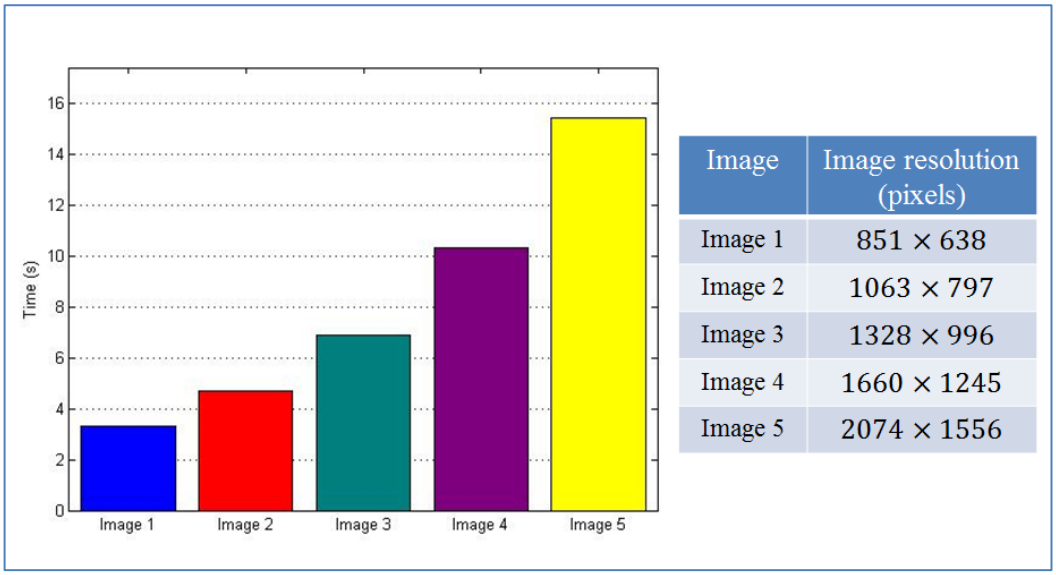


Image	Image resolution (pixels)
Image 1	851 × 638
Image 2	1063 × 797
Image 3	1328 × 996
Image 4	1660 × 1245
Image 5	2074 × 1556

Figure 5.2 : The time consumed for processing different size of images

As it can be seen the Figure 5.2, computation of time is quite dependent on resolution of image. Images with high resolution need more time. Table 5.8 presents the time consumed for different phases of algorithm applied on an image with 2074 ×1556 pixels resolution. As it can be observed, RLM metric needs more computational time compared with the other phases; whereas the classification is the fastest phase. Although RLM features needs more calculation time than other features, they are efficient features for retinal image quality assessment (see Table 5.7).

Table 5.8: Time consumed for different phases of algorithm

Algorithm phases sequence	Time (s)
Preprocessing	1.38
Sharpness + EBR	0.94
Color histogram	1.30
RLM	11.73
Classification	0.06

5.12 Alternative experiment

As second experimental procedure, 30 retinal images (15 gradable and 15 ungradable) were used for training data and 30 retinal images (15 gradable and 15 ungradable) as test data. All features (RLM, sharpness, color, EBR) were extracted and combined in a vector feature. The SVM with RBF kernel was used as classifier. The 10-fold cross validation with grid search strategy were used in order to determine optimum parameters of c, $\gamma$and n. as shown in Table 5.9

Table 5. 9: The best obtained parameters

Parameters	c	γ	n
Optimum amount	10	1	3

Table 5.10 is also showing optimum texture feature set associated to this experiment.

Table 5. 10: RLM features selected

Features selected	SRLGE	SRE	SRHGE

The accuracy, sensitivity and specificity of classification obtained from combinations of features are presented in Table 5.11

Table 5. 11 : Accuracy, sensitivity and specificity of the second experiment

features	ROI	Accuracy	sensitivity	specificity
Selected RLM, sharpness, color histogram, EBR	Optic disc and macula regions	93%	93%	93%

CONCLUSION

An automated image quality assessment algorithm was developed for classifying retinal images as gradable or ungradable. We used the no-reference image quality assessment technique because a unique reference image is not often available for image quality assessment. The proposed method relies on learning procedures to estimate image quality from training samples by measuring some generic features. The generic features include: sharpness, edge ratio, color histogram and texture. The retinal images should be sharp for a more accurate diagnosis. However, the single sharpness factor is not sufficient for diagnosing because there are a lot of sharp images that are graded as ungradable images. Hence, we used another feature called edge ratio feature. The edge ratio feature calculates the ratio of edge blocks to all blocks in each image, in order to distinguish over/under exposed images. Since sharpness and edge ratio features do not provide color information of image, color histogram features were used. Moreover, textural features were used so as to assess the information about the spatial arrangement of intensities in a retinal image. This study showed that using a large number of features might lead to overfitting, and therefore applying feature selection was necessary. Five texture features were selected as the most robust set of textural feature in order to classify input images.

One of the goals of this project was to classify images as fast as possible by avoiding structural segmentation algorithms which are time consuming. Moreover, image quality assessment procedures based on eye structural segmentation tend to be complex and are not robust, especially when dealing with low quality images.

According to the advice of ophthalmologist, the macula and optic regions are the most important regions in order diagnostic of eye diseases. Hence, we confirmed that the macula and optic disk regions are the most important regions for quality assessment. Exploiting only the macula and optic disc regions of interest, as a representative of the whole image, improved the performance of measurements in term of accuracy and time. The sensitivity and specificity were improved by 7% and 3% respectively.

Since the computational time is one of the important factors in image quality assessment, we developed a simple and fast algorithm. Moreover, ROI utilization can help to reduce the time

consumption. Also in this study, we compared two textural features, i.e. RLM and Haralik; the acquired results show that Haralik features are faster than RLM while the texture features extracted from RLM have significantly higher sensitivity than Haralik texture features. On the other hand, the accuracy of classification was improved by 9% when using RLM. Since the accuracy was more important than time consumption, the RLM features were chosen for the experiments.

The final classification led to an automated estimation of retinal images quality with high accuracy: sensitivity (94%) and specificity (92%). This indicates a very good agreement between the algorithm's prediction and medical expert's judgment.

As future work, the first aspect to follow is to test the proposed method using a larger data set in order to evaluate the performance of the method more accurately. The second aspect would be to develop an extra algorithm for automatic determination of the region of interest instead of a semi-automatic method.

REFERENCES

[1] A. A. Ventura, H. Dadgostar, and M. Taban, "Diabetic Retinopathy," in *Endocrinology and Diabetes*, ed: Springer, 2014, pp. 475-480.

[2] R. D. Jager, W. F. Mieler, and J. W. Miller, "Age-related macular degeneration," *New England Journal of Medicine*, vol. 358, pp. 2606-2617, 2008.

[3] E. T. D. R. S. R. Group, "Early photocoagulation for diabetic retinopathy: ETDRS report number 9," *Ophthalmology*, vol. 98, pp. 766-785, 1991.

[4] H. Jelinek and M. J. Cree, *Automated Image Detection of Retinal Pathology*: CRC Press, 2009.

[5] (2007). *Terese Winslow , Eye Anatomy National Cancer Institute* Available: http://www.teresewinslow.com/

[6] K. Rogers, *The eye : the physiology of human perception*: Rosen Educational Services, 2011.

[7] M. D. Abràmoff, M. K. Garvin, and M. Sonka, "Retinal imaging and image analysis," *Biomedical Engineering, IEEE Reviews in*, vol. 3, pp. 169-208, 2010.

[8] C. I. V. Center. *Eye Disease Information*. Available: http://www.centralillinoisvision.com/resources/eye_disease_information

[9] Gary H. Cassel, Michael D. Billig, Harry G. Randall, and M. F. Goldberg, *The Eye Book: A Complete Guide to Eye Disorders and Health*: Johns Hopkins University Press, 1998.

[10] (2012). *National Eye Institute* Available: https://www.flickr.com/photos/nationaleyeinstitute/

[11] S. Burk. *Optic Nerve Cupping*. Available: http://www.glaucoma.org/treatment/optic-nerve-cupping.php

[12] L. P. t. C. Blindness. *What is Age-related macular degeneration (AMD)?* Available: http://www.thelondonproject.org/OurVision/TheDisease/?id=74

[13] Ryan Prall F. (2010). *Fluorescein Angiography*. Available: http://eyewiki.aao.org/Fluorescein_Angiography

[14] J. Marc. (2010). *Optical Coherence Tomography*. Available: http://eyewiki.aao.org/Optical_Coherence_Tomography

[15] *Fundus Photography Overview*. Available: http://www.opsweb.org/?page=fundusphotography

[16] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *Image Processing, IEEE Transactions on*, vol. 13, pp. 600-612, 2004.

[17] Z. Wang, E. P. Simoncelli, and A. C. Bovik, "Multiscale structural similarity for image quality assessment," in *Signals, Systems and Computers, 2004. Conference Record of the Thirty-Seventh Asilomar Conference on*, 2003, pp. 1398-1402.

[18] Z. Wang and E. P. Simoncelli, "Reduced-reference image quality assessment using a wavelet-domain natural image statistic model," in *Electronic Imaging 2005*, 2005, pp. 149-159.

[19] R. Soundararajan and A. C. Bovik, "RRED indices: Reduced reference entropic differencing for image quality assessment," *Image Processing, IEEE Transactions on*, vol. 21, pp. 517-526, 2012.

[20] A. K. Moorthy and A. C. Bovik, "A two-step framework for constructing blind image quality indices," *Signal Processing Letters, IEEE*, vol. 17, pp. 513-516, 2010.

[21] M. A. Saad, A. C. Bovik, and C. Charrier, "A DCT statistics-based blind image quality index," *Signal Processing Letters, IEEE*, vol. 17, pp. 583-586, 2010.

[22] S. C. Lee and Y. Wang, "Automatic retinal image quality assessment and enhancement," in *Medical Imaging'99*, 1999, pp. 1581-1590.

[23] M. LalondeÝ, L. GagnonÝ, and M.-C. BoucherÞ, "Automatic visual quality assessment in optical fundus images," 2001.

[24] H. Bartling, P. Wanger, and L. Martin, "Automated quality evaluation of digital fundus photographs," *Acta ophthalmologica*, vol. 87, pp. 643-647, 2009.

[25] H. Davis, S. Russell, E. Barriga, M. Abramoff, and P. Soliz, "Vision-based, real-time retinal image quality assessment," in *Computer-Based Medical Systems, 2009. CBMS 2009. 22nd IEEE International Symposium on*, 2009, pp. 1-6.

[26] D. Usher, M. Himaga, M. Dumskyj, and J. Boyce, "Automated assessment of digital fundus image quality using detected vessel area," in *Proceedings of Medical Image Understanding and Analysis*, 2003, pp. 81-84.

[27] M. Niemeijer, M. D. Abramoff, and B. van Ginneken, "Image structure clustering for image quality verification of color retina images in diabetic retinopathy screening," *Medical image analysis*, vol. 10, pp. 888-898, 2006.

[28] A. D. Fleming, S. Philip, K. A. Goatman, J. A. Olson, and P. F. Sharp, "Automated assessment of diabetic retinal image quality based on clarity and field definition," *Investigative ophthalmology & visual science*, vol. 47, pp. 1120-1125, 2006.

[29] J. Paulus, J. Meier, R. Bock, J. Hornegger, and G. Michelson, "Automated quality assessment of retinal fundus photos," *International journal of computer assisted radiology and surgery*, vol. 5, pp. 557-564, 2010.

[30] H. Yu, C. Agurto, S. Barriga, S. C. Nemeth, P. Soliz, and G. Zamora, "Automated image quality evaluation of retinal fundus photographs in diabetic retinopathy screening," in *Image Analysis and Interpretation (SSIAI), 2012 IEEE Southwest Symposium on*, 2012, pp. 125-128.

[31] R. Ohlander, K. Price, and D. R. Reddy, "Picture segmentation using a recursive region splitting method," *Computer graphics and image processing*, vol. 8, pp. 313-333, 1978.

[32] J. Carletta, "Assessing agreement on classification tasks: the kappa statistic," *Computational linguistics*, vol. 22, pp. 249-254, 1996.

[33] K. McLaren, "XIII—The Development of the CIE 1976 (L* a* b*) Uniform Colour Space and Colour-difference Formula," *Journal of the Society of Dyers and Colourists*, vol. 92, pp. 338-341, 1976.

[34] M. D. Abramoff, B. v. Ginneken, and M. Niemeijer. *Retinopathy Online Challenge*. Available: http://webeye.ophth.uiowa.edu/ROC/var.1/www/index.php

[35] J. Lowell, A. Hunter, M. Habib, and D. Steel, "Automated quantification of fundus image quality," in *Proceedings of the 3rd European Medical and Biological Engineering Conference*, 2005, p. 1618.

[36] P. Pudil, J. Novovičová, and J. Kittler, "Floating search methods in feature selection," *Pattern recognition letters*, vol. 15, pp. 1119-1125, 1994.

[37] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern classification*: John Wiley & Sons, 2012.

[38] L. Giancardo, M. Abramoff, E. Chaum, T. Karnowski, F. Meriaudeau, and K. Tobin, "Elliptical local vessel density: a fast and robust quality metric for retinal images," in *Engineering in Medicine and Biology Society, 2008. EMBS 2008. 30th Annual International Conference of the IEEE*, 2008, pp. 3534-3537.

[39] F. Zana and J.-C. Klein, "Segmentation of vessel-like patterns using mathematical morphology and curvature evaluation," *Image Processing, IEEE Transactions on*, vol. 10, pp. 1010-1019, 2001.

[40] A. Hunter, J. A. Lowell, M. Habib, B. Ryder, A. Basu, and D. Steel, "An automated retinal image quality grading algorithm," in *Engineering in Medicine and Biology Society, EMBC, 2011 Annual International Conference of the IEEE*, 2011, pp. 5955-5958.

[41] J. Lowell, A. Hunter, D. Steel, A. Basu, R. Ryder, E. Fletcher, and L. Kennedy, "Optic nerve head segmentation," *Medical Imaging, IEEE Transactions on*, vol. 23, pp. 256-264, 2004.

[42] A. Hunter, J. Lowell, R. Ryder, A. Basu, and D. Steel, "Tram-line filtering for retinal vessel segmentation," 2005.

[43] R. M. Haralick, "Statistical and structural approaches to texture," *Proceedings of the IEEE*, vol. 67, pp. 786-804, 1979.

[44] H. Yu, S. Barriga, C. Agurto, G. Zamora, W. Bauman, and P. Soliz, "Fast vessel segmentation in retinal images using multiscale enhancement and second-order local entropy," in *SPIE medical imaging*, 2012, pp. 83151B-83151B-12.

[45] N. D. Narvekar and L. J. Karam, "A no-reference perceptual image sharpness metric based on a cumulative probability of blur detection," in *Quality of Multimedia Experience, 2009. QoMEx 2009. International Workshop on*, 2009, pp. 87-91.

[46] L. Gagnon, M. Lalonde, M. Beaulieu, and M.-C. Boucher, "Procedure to detect anatomical structures in optical fundus images," in *Medical Imaging 2001*, 2001, pp. 1218-1225.

[47] K. Goatman, A. Whitwam, A. Manivannan, J. Olson, and P. Sharp, "Colour normalisation of retinal images," in *Proceedings of Medical Image Understanding and Analysis*, 2003, pp. 49-52.

[48] F. t. Haar, "Automatic localization of the optic disc in digital colour images of the human retina," Citeseer, 2005.

[49] X. Tang, "Texture information in run-length matrices," *Image Processing, IEEE Transactions on,* vol. 7, pp. 1602-1609, 1998.

[50] P. Marziliano, F. Dufaux, S. Winkler, and T. Ebrahimi, "Perceptual blur and ringing metrics: application to JPEG2000," *Signal Processing: Image Communication,* vol. 19, pp. 163-172, 2004.

[51] E. M. A. Anas, J. G. Kim, S. Y. Lee, and M. K. Hasan, "High-quality 3D correction of ring and radiant artifacts in flat panel detector-based cone beam volume CT imaging," *Physics in medicine and biology,* vol. 56, p. 6495, 2011.

[52] J. Matousek, I. Martynov, A. Miettinen, and M. L. Infante, "Canny Edge Detector."

[53] R. M. Haralick, K. Shanmugam, and I. H. Dinstein, "Textural features for image classification," *Systems, Man and Cybernetics, IEEE Transactions on,* pp. 610-621, 1973.

[54] *HSV color space*. Available: http://www-rohan.sdsu.edu/

[55] I. Guyon and A. Elisseeff, "An introduction to variable and feature selection," *The Journal of Machine Learning Research,* vol. 3, pp. 1157-1182, 2003.

[56] Y. Saeys, I. Inza, and P. Larrañaga, "A review of feature selection techniques in bioinformatics," *bioinformatics,* vol. 23, pp. 2507-2517, 2007.

[57] X. He, D. Cai, and P. Niyogi, "Laplacian score for feature selection," in *Advances in neural information processing systems*, 2005, pp. 507-514.

[58] Richard O. Duda, Peter E. Hart, and D. G. Stork, *Pattern Classification*, Wiley-Interscience Publication ed., 2001.

[59] M. Robnik-Šikonja and I. Kononenko, "Theoretical and empirical analysis of ReliefF and RReliefF," *Machine Learning,* vol. 53, pp. 23-69, 2003.

[60] F. Nie, S. Xiang, Y. Jia, C. Zhang, and S. Yan, "Trace Ratio Criterion for Feature Selection," in *AAAI*, 2008, pp. 671-676.

[61] L. Song, A. Smola, A. Gretton, K. M. Borgwardt, and J. Bedo, "Supervised feature selection via dependence estimation," in *Proceedings of the 24th international conference on Machine learning*, 2007, pp. 823-830.

[62] Q. Gu, Z. Li, and J. Han, "Generalized fisher score for feature selection," *arXiv preprint arXiv:1202.3725*, 2012.

[63] C.-W. Hsu, C.-C. Chang, and C.-J. Lin, "A practical guide to support vector classification," ed.

[64] P. Refaeilzadeh, L. Tang, and H. Liu, "On comparison of feature selection algorithms," in *Proceedings of AAAI Workshop on Evaluation Methods for Machine Learning II*, 2007, pp. 34-39.

[65] P. Gastaldo, R. Zunino, I. Heynderickx, and E. Vicario, "Objective quality assessment of displayed images by using neural networks," *Signal Processing: Image Communication,* vol. 20, pp. 643-661, 2005.

[66] P. Gastaldo and R. Zunino, "Neural networks for the no-reference assessment of perceived quality," *Journal of Electronic Imaging,* vol. 14, pp. 033004-033004-11, 2005.

[67] P. Gastaldo, G. Parodi, J. Redi, and R. Zunino, "No-reference quality assessment of JPEG images by using CBP neural networks," in *Artificial Neural Networks–ICANN 2007*, ed: Springer, 2007, pp. 564-572.

[68] M. M. Galloway, "Texture analysis using gray level run lengths," *Computer graphics and image processing,* vol. 4, pp. 172-179, 1975.

[69] A. Chu, C. Sehgal, and J. F. Greenleaf, "Use of gray value distribution of run lengths for texture analysis," *Pattern Recognition Letters,* vol. 11, pp. 415-419, 1990.

[70] B. V. Dasarathy and E. B. Holder, "Image characterizations based on joint gray level—run length distributions," *Pattern Recognition Letters,* vol. 12, pp. 497-502, 1991.

[71] C. Cortes and V. Vapnik, "Support-vector networks," *Machine Learning,* vol. 20, pp. 273-297, 1995/09/01 1995.

[72] M. Boardman and T. Trappenberg, "A heuristic for free parameter optimization with support vector machines," in *Neural Networks, 2006. IJCNN'06. International Joint Conference on*, 2006, pp. 610-617.

[73] F. Mittag, F. Büchel, M. Saad, A. Jahn, C. Schulte, Z. Bochdanovits, J. Simón-Sánchez, M. A. Nalls, M. Keller, D. G. Hernandez, J. R. Gibbs, S. Lesage, A. Brice, P. Heutink, M. Martinez, N. W. Wood, J. Hardy, A. B. Singleton, A. Zell, T. Gasser, and M. Sharma, "Use of support vector machines for disease risk prediction in genome-wide association studies: Concerns and opportunities," *Human Mutation,* vol. 33, pp. 1708-1718, 2012.

APPENDIX A - TEXTURE ANALYSIS

Texture analysis is commonly exploited in medical image interpretation. Information in a retinal image can be deduced based on its textures. Therefore, we investigated the potential of using computerized image texture analysis to retinal image quality assessment. Efficient feature extraction is one of the most significant aspects of texture analyzing in quality assessment. In order to perform texture analysis for retinal image quality assessment the co-occurrence matrices has been widely used. Therefore, we implemented both run-length [49] and co-occurrence matrices [43] and compared their performances.

The co-occurrence matrix disclosed the spatial distribution of certain properties in the texture image. In [65-67] it was shown that features based on the co-occurrence matrix are able to provide prediction of quality.

A.1. Co-occurrence Matrix

A co-occurrence matrix, p, is an $n \times n$ matrix, where n is the number of gray-levels within images. The matrix is as an accumulator so that $p(i,j)$ counts the number of pixel pairs having the intensities i and j. Pixel pairs are defined by a distance (d) and direction (θ) which can be represented by a displacement vector $d = (dx, dy)$, where dx represents the number of pixels moved along the x-axis, and dy represents the number of pixels moved along the y-axis of the image slice. The co-occurrence matrix containing the relative frequencies $p(i,j,\theta,d)$ counting the neighborhood of each gray value to each other gray value in certain distance and direction.

$$p(i,j,0,d) = \#\{(a,x) \in [1,\dots,n],(b,y) \in [1,\dots,m] \mid f(a,b) = i, f(x,y) = j, a - x = 0, |b - y| = d\} \quad \text{(A. 1)}$$

where each p value is the number of times that: $f(x1,x2) = i, f(y1,y2) = j$, $|x1 - x2| = d$ and $y1 = y2$ append simultaneously in the image.

$p(i,j,45,d), p(i,j,90,d), p(i,j,135,d)$ are defined similarly:

$$p(i,j,45,d) = \#\{(a,x) \in [1,\dots,n],(b,y) \in [1,\dots,m] | f(a,b) = i, f(x,y) = j, (a-x = d, b-y = -d) \vee (a-x = -d, b-y = d)\} \quad \text{(A.2)}$$

$$p(i,j,90,d) = \#\{(a,x) \in [1,\dots,n],(b,y) \in [1,\dots,m] | f(a,b) = i, f(x,y) = j, |a-y| = d, b-y = 0\} \quad \text{(A.3)}$$

$$p(i,j,135,d) = \#\{(a,x) \in [1,\dots,n],(b,y) \in [1,\dots,m] | (a,b) = i, f(x,y) = j, (a-x = d, b-y = d) \vee (a-x = -d, b-y = -d)\} \quad \text{(A.4)}$$

According to the previous formulas gray-levels co-occurrence matrix depends on different parameters:

- Number of grey levels: The number of gray levels used to simplify and reduce the size of the matrix.
- Distance (d): the co-occurrence matrix stores the number of times that a certain pair of pixels is found in the image. Normally the pair of pixels is just neighbors, but the matrix could also be computed analyzing the relation between non-consecutive pixels. Thus a distance between pixels must be previously defined.
- Angle (θ): angle to match interest distance and its neighbor. The most common directions are 0, 45, 90, 135.

It is very important to decide what parameters will be used in order to obtain good features descriptor. Features derived from co-occurrence matrices are effective for describing textural properties. Haralick has proposed a number of useful texture features that can be computed from the co-occurrence matrix. Haralick features contain the relative frequencies $P(i,j)$ counting the neighborhood of each gray value to each other gray value in directions $\theta \epsilon \{0^{\circ}, 45^{\circ}, 90^{\circ}, 135^{\circ}\}$. These features are defined to emphasize texture properties statically. 13 Haralick texture features formula have been introduced in the below. In the following equations N_g refers to number of distinct gray levels in the image.

1. Energy feature is also called Uniformity feature. It measures the textural uniformity in the range. The highest energy value is for constant image

$$Energy = \sum_{i}^{N_g} \sum_{j}^{N_g} p(i,j)^2 \quad \text{(A. 5)}$$

2. Contrast feature measures of the intensity contrast between a pixel and neighbor over an image. It is the difference between the highest and the lowest values of adjacent pixels.

$$Contrast = \sum_{n=0}^{N_g-1} n^2 \left\{ \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} p(i.j) \right\}, \qquad |i-j=n| \quad \text{(A. 6)}$$

3. Correlation feature is a measure of gray tone linear dependencies on the neighboring pixels. This feature return a measure how correlated the pixel is to its neighbor whole of image.

$$Correlation = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} \frac{(i-\mu_x)(j-\mu_y)p(i,j)}{\sigma_x \sigma_y} \quad \text{(A. 7)}$$

where μ_x, μ_y, σ_x and σ_y are the means and standard deviations of p_x and p_y

4. Entropy feature measures the randomness of gray level distribution. The entropy will be maximum value when all $p(i,j)$'s are equal, or in the other word the value of this feature will be highest value when the image is not textually uniform. Entropy inversely is correlated to energy.

$$Entropy = -\sum_{i=1}^{N_g} \sum_{j=1}^{N_g} p(i,j)\ \log\{p(i,j)\} \quad \text{(A. 8)}$$

5. Variance feature measure dissimilarity of image. The value of variance feature reaches the maximum value when discrepancy between the gray level value and their mean increases.

$$Variance = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} (i - \mu)^2 \, p(i,j) \quad \text{(A. 9)}$$

6. Inverse difference moment is also called as Homogeneity. It measures image homogeneity. In other word it measures the spatial closeness of the distribution of element in the co-occurrence matrix to diagonal co-occurrence diagonal matrix. Homogeneity and contrast are correlated but inversely, it means homogeneity decreases if contrast increases while energy is kept constant.

$$Inverse\ Difference\ Moment = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} \frac{1}{1 + (i - j)^2} \, p(i,j) \quad \text{(A. 10)}$$

The rest of the Haralick features are derived from those listed above.

7. Sum Average (SA)

$$Sum\ Average = \sum_{k=2}^{2\,N_g} k p_{x+y}(k) \quad \text{(A. 11)}$$

8. Sum Entropy

$$Sum\ Entropy = -\sum_{k=2}^{2\,N_g} p_{x+y}(k) \log\{p_{x+y}(k)\} \quad \text{(A. 12)}$$

9. Sum variance

$$Sum\ Variance = \sum_{i=2}^{2N_g} (i - SA)^2 \, p_{x+y}(i) \quad \text{(A. 13)}$$

10. Difference variance

$$Difference\ Variance = \sum_{i=0}^{N_g-1} i^2 p_{x-y}(i) \quad \text{(A. 14)}$$

11. Different Entropy:

$$Different\ Entropy = -\sum_{i=0}^{N_g-1} p_{x-y}(i) \log\{p_{x-y}(i)\} \quad \text{(A. 15)}$$

12. Information Measure of Correlation 1 (IMC 1) :

$$IMC\ 1 = \frac{HXY - HXY1}{\max(HX, HY)} \quad \text{(A. 16)}$$

13. Information Measure of Correlation 2 (IMC 2) :

$$IMC\ 2 = (1 - \exp[-2(HXY2 - HXY)])^{\frac{1}{2}} \quad \text{(A. 17)}$$

where x and y are the coordinates (row and column) of an entry in the co-occurrence matrix, and $p_{x+y}(i)$ is the probability of co-occurrence matrix coordinates summing to $x + y$. HX , HY are the entropies of p_x and p_y. $HXY = -\sum_i \sum_j p(i,j) \log\{p(i,j)\}$.

$HXY1 = -\sum_i \sum_j p(i,j)\ log\{p_x(i)\ p_y(j)\}$ and $HXY2 = -\sum_i \sum_j p_x(i)\ p_y(j)\ log\{p_x(i)\ p_y(j)\}$.

A.2. Run Length Matrix

The Run Length Matrix (RLM) is another way for extracting higher order statistical texture features. Sequential pixels with the same intensity in a certain direction represent a run and run length is the number of pixels in a run. RLM extracts statistical texture features and provides a quantitative description of the features from image. The run-length matrix q is defined as follows. Each element $q(i,j)$ represents the number of runs with pixels of gray level intensity *i* and run length *j* along the directions *0 °, 45 °, 90 °, 135°*. The size of the matrix q is *m* by *n*, where *m* is the maximum gray level in the image and *n* is equal to the possible maximum run length.

Five basic features that are extracted from RLM proposed by Galloway [68] as follows: In the following equations n_r refers to the total number of runs and M is the number of gray levels. N is the number of run length. The number of runs of pixels that have gray level i and length group j is represented by $q(i,j)$.

1) Short Run Emphasis (SRE): Measures the distribution of short runs. The SRE is highly depend on the occurrence of short runs and is expected large for fine textures.

$$SRE = \frac{1}{n_r}\sum_{i=1}^{M}\sum_{j=1}^{N}\frac{q(i,j)}{j^2} \quad \text{(A. 18)}$$

2) Long Run Emphasis (LRE): Measures the distribution of long runs. The LRE is highly depend on the occurrence of long runs and is expected large for coarse textures.

$$LRE = \frac{1}{n_r}\sum_{i=1}^{M}\sum_{j=1}^{N} q(i,j) * j^2 \quad \text{(A. 19)}$$

3) Grey level Non-uniformity (GLN): Measures the similarity of the gray level value throughout the image. The GLN is low if the gray level values are similar in all over the image.

$$GLN = \frac{1}{n_r}\sum_{i=1}^{M}\left(\sum_{j=1}^{N} q(i,j)\right)^2 \quad \text{(A. 20)}$$

4) Run Length Non-uniformity (RLN): Measures the similarity of the length of runs all over the image. The RLN is expected small if the run lengths are similar all over the image.

$$RLN = \frac{1}{n_r}\sum_{j=1}^{N}\left(\sum_{i=1}^{M} q(i,j)\right)^2 \quad \text{(A. 21)}$$

5) Run Percentage (RP): Measures the homogeneity and distribution of runs of image.

$$RP = \frac{n_r}{q(i,j) * j} \quad \text{(A. 22)}$$

Chu *et al*. [69] proposed two another features in order to distinguish textures that are similar SRE and LRE features but in different gray level distribution of the runs.

6) Low Gray-Level Run Emphasis (LGRE)

$$LGRE = \frac{1}{n_r} \sum_{i=1}^{M} \sum_{j=1}^{N} \frac{q(i,j)}{i^2} \quad \text{(A. 23)}$$

7) High Gray-Level Run Emphasis (HGRE)

$$HGRE = \frac{1}{n_r} \sum_{i=1}^{M} \sum_{j=1}^{N} q(i,j) * i^2 \quad \text{(A. 24)}$$

Dasarathy *et al*.[70] were introduced four another features that are combine statistical measure of gray level and run length.

8) Short Run Low Gray-Level Emphasis (SRLGE)

$$SRLGE = \frac{1}{n_r} \sum_{i=1}^{M} \sum_{j=1}^{N} \frac{q(i,j)}{i^2 * j^2} \quad \text{(A. 25)}$$

9) Short Run High Gray-Level Emphasis (SRHGE)

$$SRHGE = \frac{1}{n_r} \sum_{i=1}^{M} \sum_{j=1}^{N} \frac{q(i,j) * i^2}{j^2} \quad \text{(A. 26)}$$

10) Long Run Low Gray-Level Emphasis (LRLGE)

$$LRLGE = \frac{1}{n_r} \sum_{i=1}^{M} \sum_{j=1}^{N} \frac{q(i,j) * j^2}{i^2} \quad \text{(A. 27)}$$

11) Long Run High Gray-Level Emphasis (LRHGE)

$$LRHGE = \frac{1}{n_r}\sum_{i=1}^{M}\sum_{j=1}^{N} q(i,j) * i^2 * j^2 \quad \text{(A. 28)}$$

APPENDIX B - CLASSIFICATION

There are a lot of classifier algorithms, such as Support Vector Machine (SVM), K Nearest Neighbors (K-NN), Naive Bayes (NB). In this study, a SVM with a radial basis function kernel was used as a classifier.

Support Vector Machine Classifier: The SVM is a popular supervised machine learning that produces a linear boundary to achieve maximum separation between two classes of subjects by kernel function of the input features for each subject. The SVM algorithm was proposed by Corinna Cortes and Vladimir Vapnik[71].

Support Vector Machine Classifier Formulation: Training data represented as vectors $\boldsymbol{x_i} \in \mathbb{R}^{\boldsymbol{p}}$ for each of *l* data points with their respective class label $\boldsymbol{y_i} \in \{-\mathbf{1}, +\mathbf{1}\}$. where $\mathbb{R}^{\boldsymbol{p}}$ is *p*-dimensional Euclidian space.

The SVM are based on the concept of linear decision function. Linear decision functions consist of decision boundary that is hyperplane separating the two different regions of the space. Figure A.1 shows in the 2D space the data separated by line (hyperplane). This line formulated by $wx + b = 0$ to separated data into two classes. The decision function f(x), assign the correct class label to each data point. SVM use a decision function of the form:

$$f(x) = sign(< w, x > + b) \quad \text{(A. 29)}$$

where <*w*,*x*> is the inner product of *w* and *x*, *w* is the hyperplan normal vector and *b* is as the bias.

The best hyperplane for an SVM means the one with the largest margin between the two classes. The vectors that define the hyperplane are the support vectors.

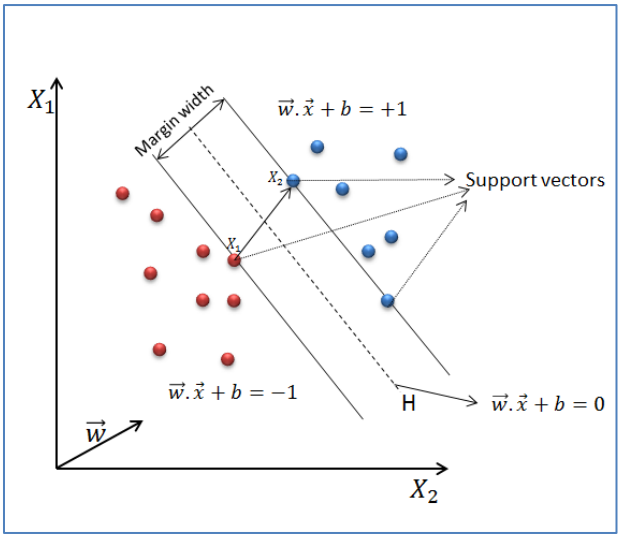


Figure A.1: A seprating hyperplan (H)

Support Vector Machine Separable Case: Figure A.1 shows a hyperplan that separate positive (blue points) and negative (red points) examples, each such hyperplane (w, *b*) is a classifier that correctly separates all patterns from the training set:

$$f(x_i)=\begin{cases}+1 & if \quad w.x_i+b>0\\ -1 & if \quad w.x_i+b<0\end{cases} \tag{A. 30}$$

For all points from the hyperplane H ($\vec{w}.\vec{x}+b=0$), the b/|w| determined the offset of the hyperplan from the origin along the normal vector w. We consider the patterns from the class -1 that satisfy the equality $\vec{w}.\vec{x}+b=-1$; and also the patterns from the class +1 satisfy the equality $\vec{w}.\vec{x}+b=+1$. The distance between hyperplanes (margin) is calculated as following:

$$\begin{aligned}\frac{w}{\|w\|}.(x_2-x_1)&=width\\ w.x_1+b&=1\\ w.x_2+b&=-1\\ w.x_1+b-w.x_2-b&=2\\ w.x_1-w.x_2&=2\\ \frac{w}{\|w\|}.(x_2-x_1)&=\frac{2}{\|w\|}\end{aligned} \tag{A. 31}$$

$$width = \frac{2}{\|w\|}$$

From these considerations it follows that the identification of the optimum separation hyperplane is performed by maximizing 2/||w||, which is equivalent to minimizing $\|w\|^2/2$. This is optimization problem. Thus, the SVM optimization problem is:

$$\min_{w,b}\left\{\frac{1}{2}\|w\|^2\right\} \; subject\ to: y_i(< w, xi > +b) \geq 1 \qquad \text{(A. 32)}$$

where $\|w\|^2 = \sqrt{< w, w >}$ is the 2-norm of w, $< w, w >$ is the dot product between w and w.

By introducing Lagrange multipliers α, the problem can be expressed as follow:

$$\min_{w,b}\max_{\alpha \geq 0}\left\{\frac{1}{2}\|w\|^2 - \sum_{i=1}^{l}\alpha_i[y_i(w.x_i - b) - 1]\right\} \qquad \text{(A. 33)}$$

Using the fact that $\|w\|^2 = w.w$ and $= \sum_{i=1}^{l}\alpha_i y_i x_i$, one can show that the problem of the SVM reduces to the following optimization problem:

$$L(\alpha) = \sum_{i=1}^{l}\alpha_i - \frac{1}{2}\sum_{i,j}\alpha_i\alpha_j y_i y_j x_i^T x_j \qquad \text{(A. 34)}$$

According to Lagrange theory, to obtain the optimum, it is enough to maximize the Lagrange with respect to α_i.

$$max.\ \sum_{i=1}^{l}\alpha_i - \frac{1}{2}\sum_{i,j}\alpha_i\alpha_j y_i y_j x_i^T x_j \;\; subject\ to: \sum_{i=1}^{l}\alpha_i y_i = 0\ , \forall i\ \alpha_i \geq 0 \qquad \text{(A. 35)}$$

Thus, the decision function f(x) calculated by equation (A. 36).

$$f(x) = sign(\sum_{i=1}^{l}\alpha_i y_i < x, x_i > +b) \qquad \text{(A. 36)}$$

Support Vector Machine Non-separable Case: The goal of SVM is produce hyperplane that completely separates the vectors into two non-overlaping classes but in some case may not be

possible classified data linearly. In this case as shown in Figure A.2, SVM find hyperplane that maximizes the margin and minimized the misclassifications (soft-margin method). This method uses slack variable $\boldsymbol{\xi_i}$, which allow data to violate the margin.

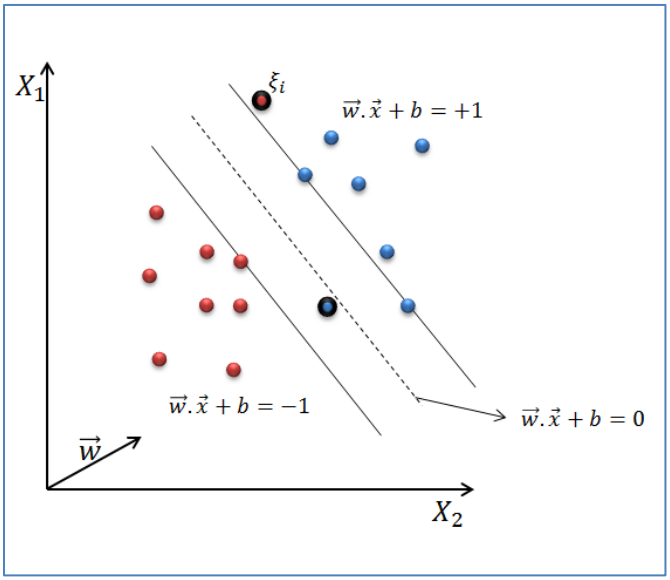


Figure A.2 : Maximum-margin hyperplane and minimum misclassification

The penalty for this violation is controlled by the cost parameter *C*. The soft margin SVM optimization can be then defined as:

$$\min_{w,b,\xi}\left\{\frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n}\xi_i\right\} \; subject\ to: \quad y_i(< w, xi > +b) \geq 1 - \xi_i \qquad \xi_i \geq 0. \qquad \text{(A. 37)}$$

Variable *C* is trades-off margin width and misclassifications. The choice of *C* is important step of classification because it directly affects the generalization ability of the learned models.

By using Lagrange multipliers α, the problem of optimization can be expressed as follow:

$$\min_{w,b,\xi}\max_{\alpha,\beta}\left\{\frac{1}{2}\|w\|^2 - C\sum_{i=1}^{l}\xi_i - \sum_{i=1}^{l}\alpha_i[y_i(w.x_i - b) - 1 + \xi_i] - \sum_{i=1}^{l}\beta_i\xi_i\right\} \qquad \text{(A. 38)}$$

with $\alpha_i , \beta_i > 0$.

Kernel Trick: Some binary classification problems do not have a simple hyperplane as a useful separating criterion. SVM handles this by applying the kernel function ($\boldsymbol{K(x_i, x_j)}$) to map data into a higher dimensional space by the function $\boldsymbol{\phi}$ to make it possible to perform the linear separation (Figure A.3). Furthermore $\boldsymbol{K(x_i, x_j) \equiv \phi(x_i)^T \phi(x_j)}$ is called kernel function. Thus, the decision function reformulated as Equation (A. 39).

$$f(x) = sign(\sum_{i=1}^{l} \alpha_i y_i K(x, x_i) + b \qquad \text{(A. 39)}$$

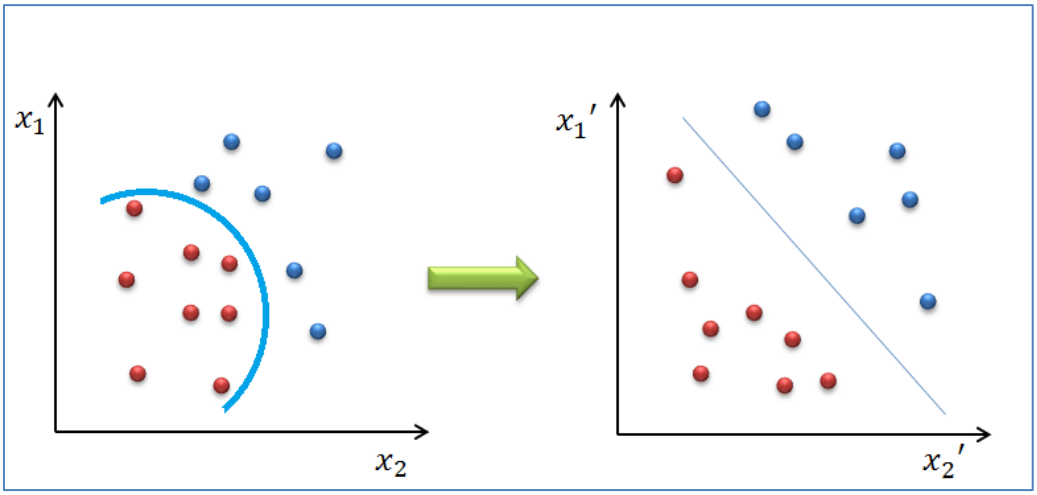


Figure A.3 : Mapping data onto another space in order to classify data linearly

Some common Kernel functions include the linear kernel, the polynomial kernel and the Gaussian kernel. Below is a simple list with their formula.

- The linear kernel: $K(x_i, x_j) = x_i^T x_j$.
- The Gaussian kernel: $K(x_i, x_j) = (\gamma x_i^T x_j + r)^d, \ \gamma > 0.$
- The polynomial kernel: $K(x_i, x_j) = \exp\left(-\gamma \left\| x_i - x_j \right\|^2\right), \gamma > 0.$

Here γ, r and d are kernel parameters.

The simplest kernel function is linear kernel. It is given by the common inner product <x,y> plus an optional constant c. The SVM using a linear kernel are often equivalent to their non-kernel. So, this kernel cannot handle the case may not be possible classified data linearly. The Polynomial kernel is a non-stationary kernel. It is well suited for problems where all data are normalized. The number of hyperparameters influences the complexity of model selection[63]. The polynomial kernel has more parameter than Gaussian kernel. The Gaussian kernel is known as radial basis function (RBF), RBF is one of the most popular kernel functions with convincing performance and is a reasonable to choice.

Cross-Validation:

Cross-Validation (CV) is a statistic method of evaluating algorithm, by dividing data into two parts: one used for train a model and other part use for validate the model.

K-fold Cross-Validation: The basic form of CV is k-fold cross-validation. In k-fold cross-validation, data divided into k equal parts or folds. One fold used for validation and k-1 folds used for training data which training and validation data are performed k times. Final result is the average value obtained from the accuracies of the k iteration. Figure A. 4 shows an example with k=5.

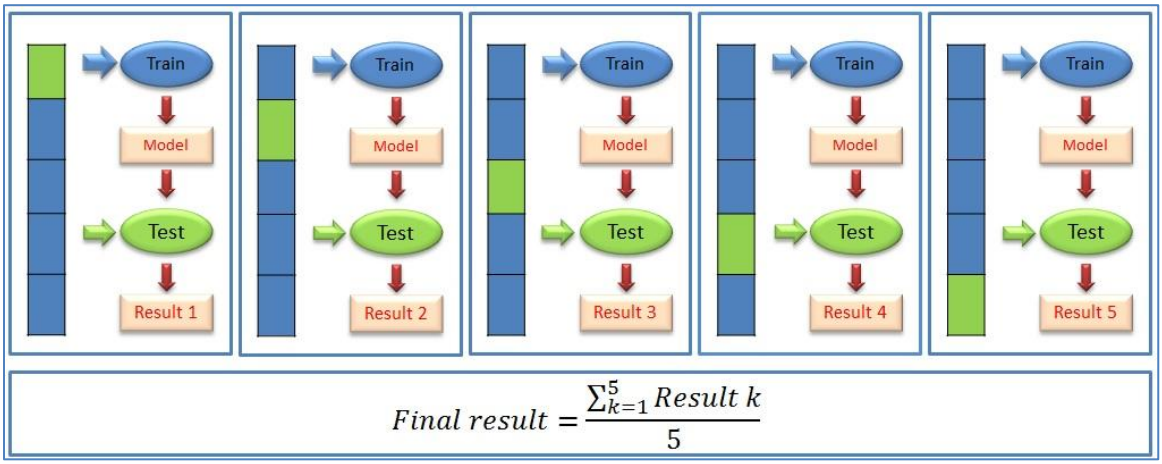


Figure A. 4: Procedure of 5-fold cross validation

Leave-one-out cross-validation: Leave-one-out cross-validation (LOOCV) is the same as k-fold cross validation where k equals the number of whole data. Hence, in the each iteration of LOOCV one single data used as validate data and remaining data used for training.

There are two goals for using CV:

- Performance estimation
- Adjustment training model parameters.

Performance estimation: The SVM classification task involves separating data into training and testing sets. A common strategy is to segment dataset into two segments: train and test data. This procedure avoids the overlap between training data and test set but the results are highly dependent on the choice for the training or testing segment. To deal with these challenges and utilize the available data to the as much as possible, CV is used. In other words, CV allows for all data to be used to estimate performance of algorithm. The Accuracy indicator is used to evaluate the performance of algorithm. Accuracy is defined as:

$$Accuracy = \frac{tp + tn}{tp + fp + fn + tn} \tag{A. 40}$$

where tp- true positive, fp- false positive, fn-false negative and tn- true negative counts. The accuracy indicator evaluates how a certain classifier predict label of test data.

*Adjustment training model parameters***:** A Support Vector Machine (SVM) with a Radial Basis Function (RBF) kernel was used as a classifier. There are two parameters for an RBF kernel: *C* and γ. It is not known beforehand which *C* and γ are best for a given problem. The accuracy of an SVM model depends on the selection of *C* and γ.

The parameter *C* controls the trade-off between maximization of the margin width and minimizing the number of misclassified points in the training set[72]. In other words choice of C directly affects the generalization ability of learned model. Increasing the value of *C* increases penalty for misclassifying points that leads to overfitting. Conversely selecting small *C* value increases the margin that may lead to misclassified points in the training set. Figure A.5 shows an example that demonstrates the dependency of the soft-margin SVM on different *C* values. Although the data is linearly separable, when *C* is small (left panel), the margin increased and there are 2 misclassification that leads to failure in learning. When the *C* value is increased (right

panel), a narrow margin is obtained. The resulting decision boundary leads to correct classification of the training data, but the small margin creates a more accurate model that may not generalize well.

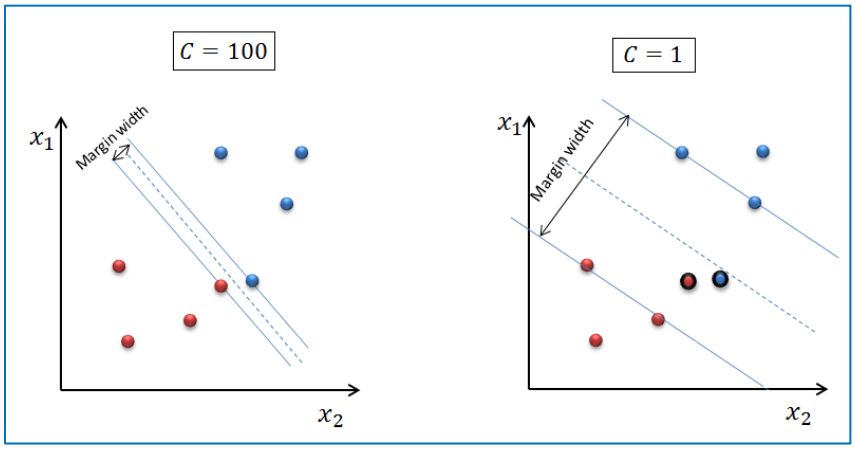


Figure A.5: Effect of SVM parameter (C)

For selecting small Gaussian parameter (γ) the decision boundary is nearly linear. When γ value increased the flexibility of the decision boundary increases that leads to overfitting [73].